Assessing cognitive presence using automated learning analytics methods

Vitomir Kovanović

Doctor of Philosophy

Institute for Language, Cognition and Computation
School of Informatics
University of Edinburgh

2017

Vitomir Kovanović:

Assessing cognitive presence using automated learning analytics methods

Doctor of Philosophy, 2017

PRINCIPAL SUPERVISOR:

Professor Dragan Gašević, Ph.D., *The University of Edinburgh, United Kingdom.*

ASSISTANT SUPERVISORS:

Professor Marek Hatala, Ph.D., *Simon Fraser University, Canada.*

Professor George Siemens, Ph.D., *The University of Texas at Arlington, United States.*

EXAMINATION COMMITTEE:

Professor Helen Pain, Ph.D., *The University of Edinburgh, United Kingdom.*

Professor Terry Anderson, Ph.D., *Athabasca University, Canada.*

Posvećeno Dunji i Jeleni.

Abstract

With the increasing pace of technological changes in the modern society, there has been a growing interest from educators, business leaders, and policymakers in teaching important higher-order skills which were identified as necessary for thriving in the present-day globalized economy. In this regard, one of the most widely discussed higher order skills is critical thinking, whose importance in shaping problem solving, decision making, and logical thinking has been recognized. Within the domain of distance and online education, the Community of Inquiry (CoI) model provides a pedagogical framework for understanding the critical dimensions of student learning and factors which impact the development of student critical thinking. The CoI model follows the social-constructivist perspective on learning in which learning is seen as happening in both individual minds of learners and through the discourse within the group of learners. Central to the CoI model is the construct of cognitive presence, which captures the student cognitive engagement and the development of critical thinking and deep thinking skills. However, the assessment of cognitive presence is challenging task, particularly given its latent nature and the inherent physical and time separation between students and instructors in distance education settings. One way to address this problem is to make use of the vast amounts of learning data being collected by learning systems.

This thesis presents novel methods for understanding and assessing the levels of cognitive presence based on learning analytics techniques and the data collected by learning environments. We first outline a comprehensive model for cognitive presence assessment which builds on the well-established evidence-cantered design (ECD) assessment framework. The proposed assessment model provides a foundation of the thesis, showing how the developed analytical models and their components fit together and how they can be adjusted for new learning contexts. The thesis shows two distinct and complementary analytical methods for assessing students' cognitive presence and its development. The first method is based on the automated classification of student discussion messages and captures learning as it is observed in the student dialogue. The second analytics method relies on the analysis of log data of students' use of the learning platform and captures the individual dimension of the learning process. The developed analytics also extend current theoretical understanding of the cognitive presence construct through data-informed operationalization of cognitive presence with different quantitative measures extracted from the student use of online discussions.

We also examine methodological challenges of assessing cognitive presence and other forms of cognitive engagement through the analysis of trace data. Finally, with the intent of enabling for the wider adoption of the CoI model for new online learning modalities, the last two chapters examine the use of developed analytics within the context of Massive Open Online Courses (MOOCs). Given the substantial differences between traditional online and MOOC contexts, we first evaluate the suitability of the CoI model for MOOC settings and then assess students' cognitive presence using the data collected by the MOOC platform. We conclude the thesis with the discussion of practical application and impact of the present work and the directions for the future research.

Lay summary

This thesis presents novel methods for assessing students' cognitive presence, a theoretical construct which captures development of students' critical and deep thinking skills within distance learning setting. By the means of learning analysis methods, we show how the data collected by the learning systems can be used to assess and understand students' cognitive presence, which is shown to directly affect their logical thinking, decision making, and problem solving skills. In the thesis, we first provide overview of the conceptual model for cognitive presence assessment, which lays the foundation for the thesis and shows how different elements of the models are combined to provide comprehensive understanding of student cognitive presence. We next describe two complementary analytical models for assessing cognitive presence using different types of learning data. Finally, we examine the use of the analytics model developed within the context of Massive Open Online Courses and show how the differences between traditional online courses and MOOC affect student cognitive presence development.

Acknowledgment

This thesis is a result of a challenging, but profoundly life-changing journey. Here I would like to thank several individuals for their kind help and support without which this thesis would not have been possible.

First and foremost, I would like to express the deepest gratitude to my supervisor, Professor Dragan Gašević for his enthusiasm, patience, and immense knowledge without which this thesis would not have materialized. His help and support far exceeded his formal supervisory role and enabled me to grow both academically and personally. I cannot express how much honored and tremendously grateful I am for having Professor Gašević as my supervisor and mentor, he is what any student can only wish for.

I also want to thank my assistant supervisors, Professor Marek Hatala and Professor George Siemens, whose constructive feedback, comments, and words of encouragement were essential for the work presented in this thesis. I would also like to offer special thanks to my closest colleague and dear friend, Srećko Joksimović, who was together with me on this long journey through the graduate school. His numerous comments, insights, and ideas were invaluable for this thesis, and without him, it would not have been possible.

During my doctoral studies, I was fortunate enough to work with an incredible group of friends and colleagues who made this journey so much more enjoyable. I would particularly like to thank Oleksandra Poquet, Iva Ćukić, Zak Waters, Thieme Hennis, Kirsty Kitto, Professor Shane Dawson, Professor Ryan Baker, Professor Pieter de Vries, and Professor Olusola Adesope for their valuable contributions to the work presented in this thesis.

I would also like to thank my brother Miroslav Kovanovic and my sister-in-law Marija Kovanovic for their unconditional love and support during my childhood years. They were the first to show me the importance of hard work and learning which made me the man I am today. Finally, I want to express the deepest appreciation to my wife Jelena Kovanovic, for her love and understanding during the last six years, and my beloved daughter Dunja Kovanovic, for brightening my days and making me see the world in a different light. They were the ones comforting me at the most difficult times, never doubting me when all seemed dark, and for that, I am eternally grateful. Thank you.

Declaration of authorship

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own. This thesis also includes six peer-reviewed publications produced under the joint authorship:

(1) Kovanović, V., Gašević, D., Hatala, M., & Siemens, G. (2017). *A novel model of cognitive presence assessment using automated learning analytics methods*. SRI Education. Retrieved from http://a4li.sri.com/archive/papers/Kovanovic_2017_Presence.pdf

(2) Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK'16)* (pp. 15–24). LAK '16. New York, NY, USA: ACM. doi:10.1145/2883851.2883950

(3) Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, *27*, 74–89. doi:10.1016/j.iheduc.2015.06.002

(4) Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., & Baker, R. S. (2015). Does time-on-task estimation matter? Implications on validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. doi:10.18608/jla.2015.23.6

(5) Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., Čukić, I., de Vries, P., Hatala, M., Dawson, S., Siemens, G., & Gašević, D. (2017). Exploring communities of inquiry in massive open online courses. *Manuscript submitted for publication*

(6) Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., de Vries, P., Hatala, M., Dawson, S., Siemens, G., & Gašević, D. (2017). The role of technology use on shaping student learning experience in MOOCs. *Manuscript submitted for publication*

I declare that I substantially contributed to all six publications (i.e., over 50% of the work done) and was involved in all phases of the research process, including study conceptualization, data collection, data analysis and interpretation, as well as the writing of the final publication.

Vitomir Kovanović

Vitomir Kovanović

Table of contents

1 Introduction

It is a mistake to think you can solve any major problems just with potatoes.

— Douglas Adams, *Life, the Universe, and Everything*

While technology has always been an integral part of teaching and learning, recent developments of information technologies and abundance of digital data brought substantial changes to the educational landscape. The recent 2017 NMC Horizon Report (Adams Becker et al., 2017) identified online, mobile, and blended learning as one of the ten critical components of higher education, stating that *"if institutions do not already have robust strategies for integrating these now pervasive approaches, then they simply will not survive"* (Adams Becker et al., 2017, p. 2). The same report also notices the high potential of big data and analytics for driving the evidence-based educational innovation and continuous monitoring of student learning (Adams Becker et al., 2017).

The rapid technological advancements also resulted in the increased interest in the development of students' critical and deep thinking skills. The significance of critical thinking has long been recognized (Dewey, 1910), and its role in shaping logical thinking, decision making, and problem-solving (Liu, Frankel, & Roohr, 2014). Over the years, critical thinking received growing interest from educators, business leaders, and policymakers, with the increasing number of institutions including it as one of its core learning outcomes (Liu et al., 2014). Together with creativity, communication, and collaboration, critical thinking has been labeled one of the core *21st-century skills* and recognized as vital for thriving in the modern day global economy (Trilling & Fadel, 2009).

Within the context of education, the importance of social interactions for the development of student critical thinking has long been recognized (Lipman, 1991). With this in mind, one pedagogical approach specifically designed to promote students' development of critical thinking by means of social interactions is inquiry-based learning. Instead of presenting students with already established facts and information, inquiry-based learning starts with a question, problem, or scenario which is designed to trigger a learning cycle in which students develop new knowledge by sustained interaction with learning materials, peer learners, and instructors. In the context of online and distance education, the *Community of Inquiry (CoI) model* (Figure 1) by Garrison, Anderson, and Archer

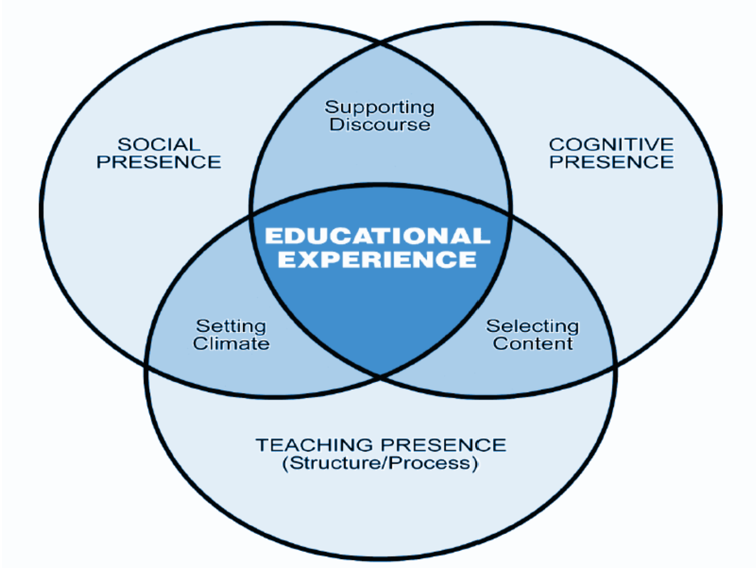


Figure 1. Community of Inquiry (CoI) model (adopted from Garrison et al., 2001)

(1999) is arguably the most widely-used and well-developed pedagogical framework (Jézégou, 2010) which focuses on the development of students' critical and deep thinking skills using inquiry-based learning through student sustained communication (Shea & Bidjerano, 2010). The CoI model defines three dimensions of online learning – also known as presences – which together shape student online educational experience (Garrison et al., 1999): 1) *cognitive presence*, which captures students' development of critical and deep thinking skills (Garrison, Anderson, & Archer, 2001); 2) *social presence*, which represents student interactions and social climate of the course (Rourke, Anderson, Garrison, & Archer, 1999); and 3) *teaching presence*, which captures instructional role before and during the course (Anderson, Rourke, Garrison, & Archer, 2001).

Due to their latent nature and the inherent separation between learners and instructors within the distance education context, the assessment of three CoI presences is a challenging task (Akyol & Garrison, 2011b). Hence, the CoI model also provides an instrument for assessing students' levels of three presences through quantitative content analysis (Krippendorff, 2003) of student discussions, which are in most instances the primary means of social interactions among students and instructors. Given the significant methodological challenges associated with the transcript analysis (Rourke, Anderson, Garrison, & Archer, 2001), as well as substantial time and effort involved, a self-reported survey instrument for assessing the perceived levels of three presences has also been developed (Arbaugh et al., 2008). However, neither of those two methods provide an effective way of assessing students' development of cognitive presence in a continuous manner that can facilitate instructional interventions in real-time. Hence, the primary use of both instruments has been for the posthoc analysis of the student learning, primarily for research purposes (Kovanović, Gašević, & Hatala, 2014). One way to address this issue is to make use of the vast amounts of data available within online learning environments for the development of analytical systems for assessing students as they progress through the courses, and that is the central idea of this thesis.

In this PhD thesis, we present novel methods for assessing and understanding students cognitive presence development and their engagement in online learning contexts. Through the analysis of the different types of data collected by online learning environments, this thesis presents two distinct and compatible analytical models for assessing students' cognitive presence. The first model makes use of the student discussion transcripts and provides means of assessing students' cognitive presence as it develops through online discourse. The second model provides insights into the personal dimension of student learning through the analysis of trace data about students' use of a learning platform. Taken together, those two models provide a comprehensive and holistic view of student cognitive presence development and, more broadly, student course engagement.

1.1 Research goals and questions

This work presented in this thesis was conducted with four primary research goals in mind. The first goal is to develop an assessment model of cognitive presence construct, which operationalizes critical elements of cognitive presence assessment systems. The model can be then used to design and develop theoretically sound learning analytics systems which are particularly tailored to the different learning contexts and scenarios. Thus, our first research question is

RESEARCH QUESTION 1:

How can we develop a flexible and theoretically sound model of cognitive presence assessment, which is both detailed enough to provide practical guidance for the development of analytics models, and, at the same time, flexible enough to accommodate assessment of cognitive presence within various learning contexts?

The second goal of the present thesis is to provide novel insights into the nature of cognitive presence, that will improve the theoretical understanding of cognitive presence construct and in more general terms, inquiry-based learning. Using the automated learning analytics methods and techniques, the goal is to examine what are the characteristics of different phases of cognitive presence, as expresses in the discussion transcripts. As such, our second research questions is

RESEARCH QUESTION 2:

How are different phases of cognitive presence expressed in the discussion transcripts? What quantitative measures extracted from the student discussion messages are reflective of the different phases of cognitive presence?

The third goal of the present research is to design and develop automated learning analytics systems using the data collected by the learning systems. In particular, through the analysis of student discussion transcripts and trace data about students' interactions with the learning platform, we aim to provide holistic assessment of students' cognitive presence. Looking specifically at cognitive presence assessment within traditional, for-credit online courses, the goal is to provide easy to use, fast and reliable method for assessing students' cognitive presence. The development of such analytics system would enable broader adoption of the CoI model by researchers and practitioners, which is currently challenging using the existing assessment techniques of (Garrison et al., 2001) and Arbaugh et al. (2008). With this in mind, our third research question is

Research Question 3:

To what extent can we use learning analytics to assess student levels of cognitive presence? How accurately can an automated analytics system gauge the levels of students' cognitive presence and its development through automated analysis of discussion transcripts and trace data of students' interactions with learning platform?

Finally, with the recent developments of new modalities of online learning, such as massive open online courses (MOOCs), the fourth goal of this thesis is to examine whether it is possible to provide assessment of students' cognitive presence within those new online learning modalities. Given the much larger and diverse student cohorts in those courses and instructors' limited capacity, the development of such system can information for instructors on students' development of cognitive presence, and guide their attention to the places where it is mostly needed. As such, our fourth research question is

Research Question 4:

Can we enable assessment of cognitive presence within newer modalities of online learning, such as MOOCs? How does cognitive presence development within MOOCs differ from traditional online courses?

However, given the important differences between traditional, small scale online courses and MOOCs regarding course organization and design, before we can address RQ 4, we first need to examine whether the use of the CoI model within the MOOC context is theoretically justified and valid. As such, we need to examine whether the differences between traditional small scale online courses and MOOC negatively affect the validity of the model for assessing students' learning experience. As such, the fifth and final research question of this thesis is:

Research Question 5:

Can we use the CoI model to assess student learning experience in MOOCs? Are there any important differences with respect to the development of three presences between MOOCs and traditional online courses? Are there any important differences with respect to the cognitive presence development between MOOC and traditional online courses?

1.2 Methodology

The primary means of conducting the research in this thesis are quantitative learning analytics methods and the investigation of empirical data from real-world, traditional online and MOOC courses.

First, we utilized Evidence-Centered Design (ECD) framework (Mislevy, Almond, & Lukas, 2003) for the development of an assessment model for cognitive presence which outlines the components of analytical models and how they fit together into a coherent unit (RQ1). Next, we used text mining and text classification methods (Aggarwal & Zhai, 2012) for the extraction of the quantitative metrics from students' discussion transcripts. The extracted measures were then used to novel insights into the cognitive presence and to provide data-informed characterization of its phases (RQ2), and also for the development of a classification system that categorizes discussion messages according to their cognitive presence (RQ3). To collect the training data for our classifier we used quantitative content analysis (QCA) (Krippendorff, 2003) of student discussion messages using the CoI coding schemes for cognitive presence defined by Garrison et al. (2001). In a similar manner, unsupervised data mining techniques were used to assess student interactions and engagement with the learning platform, which – together with the assessment of student discourse – provide a holistic assessment of the students' cognitive presence development (RQ3). Finally, before developing a learning analytics model for assessing cognitive presence within MOOC settings (RQ4), we used Exploratory Factor Analysis (EFA) (Field, Miles, & Field, 2012) to evaluate the differences between traditional, for-credit online courses and MOOCs with regards to the CoI model.

1.3 Thesis structure and overview

To address the four research questions in a methodologically rigorous and sound manner, we focused our efforts on several related problem domains, organized as six individual thesis chapters as shown in Figure 2. The overall thesis is structured across three learning analytics development stages of conceptualization, design, and implementation (Figure 2), with different chapters corresponding to one or more stages. Each chapter focuses on one or more research questions (Table 1), and includes one peer-reviewed publication which constitutes the core of the chapter. We also provide a short preface and summary to each included publication to describe how a particular publication fits into the overall structure and topic of the thesis.

In the remainder of this section, we provide a brief overview of each chapter and how they contribute to the overall research goal of the thesis.

1.3.1 Overview of chapter two: "Cognitive presence assessment model" (RQ 1)

In order to provide a coherent and comprehensive assessment of cognitive presence in a methodologically sound manner, we first develop a model for cognitive presence assessment. The model is based on the widely-used evidence-centered design (ECD) (Mislevy et al., 2003) model of educational assessment design. The ECD framework lays the foundation for the work presented in this thesis, outlining how the learning analytics models and their components fit together to provide a coherent and holistic assessment of students' cognitive presence development.

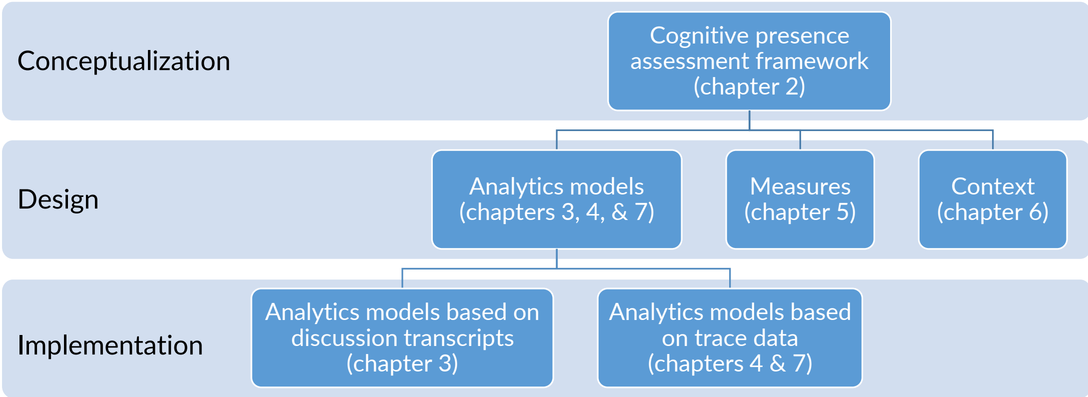


Figure 2. Graphical structure of the thesis

Research contributions:

- We developed an assessment model for cognitive presence using learning analytics methods.
- The model provides a conceptual framework for designing, implementing, and customizing the analytics implementations for particular learning contexts.
- The model shows how different types of analytics can be used to provide a coherent and comprehensive view of student cognitive presence.

Research output:

1. Kovanović, V., Gašević, D., Hatala, M., and Siemens, G. (2017): *"A novel model of cognitive presence assessment using automated learning analytics methods"*: A report article describing the cognitive presence assessment model published by the SRI International as a part of Analytics4Learning report series [1].

Table 1. Overview of the thesis research questions by individual chapters.

		Research questions				
Chapter	**Title**	RQ 1	RQ 2	RQ 3	RQ 4	RQ 5
Chapter 2	Cognitive presence assessment model	✓				
Chapter 3	Assessing cognitive presence from online discussion transcripts		✓	✓		
Chapter 4	Assessing cognitive presence using trace data			✓		
Chapter 5	Methodological challenges with trace data usage for learning analytics models			✓		
Chapter 6	Validation of the Community of Inquiry model within MOOC context					✓
Chapter 7	Assessing cognitive presence within MOOC courses				✓	

[1] http://improvement-analytics.org/a4l-network

1.3.2 Overview of chapter three: "Assessing cognitive presence from online discussion transcripts" (RQs 2 & 3)

Given that social interactions and sustained dialogue represent key learning processes through which students develop their cognitive presence, the first analytical model is based on the analysis of student discussion transcripts. In this chapter, we describe a text classification model for categorizing discussion messages based on their cognitive presence *phase*, which captures different phases of inquiry-based learning cycle (Garrison & Archer, 2000). The model is based on various types of quantitative measures extracted from the structure (e.g., the position of the message within discussion) and content (e.g., the number of words in the message, the message cohesion, number of content words) of student discourse. To train our classification system, we used the dataset of manually coded student discussion with the CoI coding scheme for cognitive presence (Garrison et al., 2001). The training data were used to automatically discover underlying rules and patterns in the data, which can be then used to categorize new, previously not seen, data points.

Research contributions:

- We developed a discussion message classification system which can be used to classify student discussion messages based on their level of cognitive presence.
- The developed classification scheme obtained inter-rater reliability of Cohen's $\kappa = 0.63$ which is significantly higher than what is reported in the existing systems for cognitive presence classification (McKlin, 2004; Corich, Hunt, & Hunt, 2012).
- We identified data-informed indicators of different phases of cognitive presence in student discussion transcripts which provide additional insights into the nature of cognitive presence and the phases of inquiry-based learning.

Research output:

1. Kovanović, V., Joksimović, S., Gašević, D., and Hatala, M. (2014): *"Automated Cognitive Presence Detection in Online Discussion Transcripts"*: A workshop paper describing the development of the first version of text classification system, based on support vector machine (SVM) algorithm and presented at the Learning Analytics and Machine Learning (LAML) Workshop at the Fourth International Learning Analytics and Knowledge Conference *(LAK'14)*.
2. Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., and Siemens, G. (2016): *"Towards automated content analysis of discussion transcripts: A cognitive presence case"*: A full conference paper describing the development of the final version of text classification system, based on Random Forests (RF) and presented at the Sixth International Learning Analytics and Knowledge Conference *(LAK'16)*.

1.3.3 Overview of chapter four: "Assessing cognitive presence using trace data" (RQ 3)

As indicated by Garrison et al. (2001), within inquiry-based learning, students constantly move between the shared world of discourse and the private, individual world of reflective thinking. As such, in addition to discourse-based assessment of the student cognitive presence development, it is equally important to assess students' critical thinking and cognitive engagement that take place outside the discourse. While it is not possible to examine what is happening inside the minds of individual learners, learning environments collect vast amounts of data (often called trace data, log data, or clickstream data) relating to students' use of learning platform and navigation through digital tools and resources, which can be used to better understand student cognitive presence and course engagement. The focus of this chapter is on examining students' cognitive presence through the assessment of students' use of available educational tools and resources available within the learning environments. In particular, through a cluster analysis of trace data about student interaction with(in) a learning environment, we identified different student learning strategies based on their technology use and then examined what are the key differences between the identified learning strategies with respect to their cognitive presence development.

Research contributions:

- We developed a method for identification of different learning strategies with respect to students' use of tools and resources available in the online learning environments.
- Using the data from several offers of a fully-online for-credit course, we identified different learning strategies and examined their association with the difference in cognitive presence development.
- Our results reveal several ways in which students in online courses can be successful and indicate the need for specific instructional interventions to cater to the needs of students with different learning strategies.

Research output:

1. Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., and Adesope, O. (2015): *"Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions"*: A journal article describing the identified student learning strategies and their association with the different levels of student cognitive presence and course learning outcomes, published in the Internet and Higher Education journal.

1.3.4 Overview of chapter five: "Methodological challenges with trace data usage for learning analytics models" (RQ 3)

As assessment of cognitive presence using trace data involves measuring student engagement with various online tools and resources, it is essential that student engagement is accurately estimated. A popular approach to estimating student engagement is to gauge time spent on different learning activities, which is commonly known as student *time-on-task* (Carroll, 1963; Bloom, 1974). However, as most web-based learning systems only record times when important events happened (e.g., at time `t1` student `s1` accessed resource `r1`, at time `t2` student `s1` accessed discussion `d1`), student time-on-task must be reconstructed from the trace data. This reconstruction processes poses a significant number of methodological challenges. The focus of this chapter is on reviewing the ways in which student time-on-task is estimated within the learning analytics field, and also what are the effects of the different estimation approaches on the accuracy of assessment of cognitive presence and other learning outcomes.

Research contributions:

- We provide an overview of the different methods for estimation of students' time-on-task that were used in the learning analytics literature.
- We examined the effects of the adopted estimation method on the results of statistical analysis within traditional for-credit online courses. In particular, we investigated the effect of the adopted time-on-task estimation method on the relationship between student time-on-task and several course outcomes (i.e., the level of cognitive presence, discussion participation grade, assignment grades, and final course grade).
- We also replicated the analysis within the blended online courses, to examine the role of particular learning modality on the importance of time-on-task estimation strategies. In particular, we explored the role of time-on-task estimation on the relationship between extracted count and time-on-task measures and final course grade within the context of 9 different blended courses.
- Our results revealed that a significant portion of variance explained of around 15% in the examined regression models can be attributed solely to adopted time-on-task estimation method, which indicates significant methodological challenges with the use of time-on-task measures of student engagement.

Research output:

1. Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., and Hatala, M. (2015): *"Penetrating the black box of time-on-task estimation"*: A full conference paper describing the implications of the different approaches for calculating student time-on-task measures and their effect on the results of the analytical models. The paper was presented at the Fifth International Conference on Learning Analytics and Knowledge *(LAK'15)* and was awarded the best paper award.
2. Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., and Baker, R. S. (2015): *"Does time-on-task estimation matter? Implications on validity of learning analytics findings"*: A journal article published by the Journal of Learning analytics which is an extension of the Kovanović, Gašević, Dawson, Joksimović, Baker, and Hatala (2015) study. The article provides more comprehensive analysis of different methods for time-on-task estimation on analytics model results within both traditional online and blended courses.

1.3.5 Overview of chapter six: "Validation of the Community of Inquiry model within MOOC context" (RQ 5)

The critical component of the CoI model, like all social-constructivist learning models, is that it presupposes sustained social interactions among students which are facilitated by a strong instructor's presence in a course (Garrison, 1993; Garrison et al., 1999). Given the high demand for the instructor's participation, courses that adopt the CoI and similar models are in practice rarely used for courses with more than 30-40 students (Anderson & Dron, 2010). With the development of analytical models – such as the ones presented in this thesis – there is a potential for scaling up the existing pedagogical models of online learning beyond traditional small-scale, for-credit online courses, to newer massive modalities of online learning – such as MOOCs – where it is not feasible for instructors to attend to every student individually. However, before different analytics models for assessing cognitive presence can be applied within a new setting in a pedagogically sound manner, it is important to first investigate the suitability of the CoI model for understanding student learning experiences in MOOC courses. The focus of this chapter is to test the adequacy of the CoI model withing MOOC context. To answer this research question (RQ 5), we examined and validated student responses to the CoI survey instrument (Arbaugh et al., 2008) and whether the factor structure of the survey latent constructs that represent three presences is preserved.

Research contributions:

- We investigated the validity of the CoI model within the context of MOOCs. More precisely, we examined the factor structure and reliability of the CoI survey instrument using the data of 1,887 students from five MOOCs.
- We examined the fit of the original three-factor model of the CoI survey instrument and also the optimal factor structure.
- Our results confirm the reliability and validity of the CoI instrument for assessing student learning experience in MOOCs.
- The investigation also reveals significant differences between MOOCs and traditional online courses with respect to the student perceptions of the three CoI presences.

Research output:

1. Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., Čukić, I., de Vries, P., Hatala, M., Dawson, S., Siemens, G., and Gašević, D. (2017): *"Exploring communities of inquiry in massive open online courses"*: A journal article on the validation of the CoI model within MOOC courses, currently under the review.

1.3.6 Overview of chapter seven: "Assessing cognitive presence within MOOC courses" (RQ 4)

Following the validation of the CoI model within the MOOC setting, the penultimate chapter of the thesis provides an implementation of an analytical model that can be used for assessing students' cognitive presence development within MOOCs. Using the CoI survey instrument (Arbaugh et al., 2008) and student cluster analysis based on their trace data records collected by the MOOC platform. Our results indicate that the method developed originally for the traditional, small-scale online courses can be successfully adjusted for the use within the MOOC context to provide insights into the development of cognitive presence. Given the challenges of assessing students' cognitive presence and course engagement in large student cohorts, this study provides the first step towards a pedagogically sound and theory-driven use of trace data about student interactions with(in) a MOOC platform as a source of insights into student learning processes.

Research contributions:

- We developed a method for identification of different learning strategies based on students' technology use within the MOOC courses. The model developed in this chapter builds extensively on the model by Kovanović, Gašević, Joksimović, et al. (2015) introduced in Chapter four and adapts it to account for the differences between MOOC and traditional online courses.
- Through the analysis of data about 23,648 students, we identified three different learning strategies within MOOC courses and examined their relationship with the perceived levels of cognitive, social, and teaching presence. We also investigated the relationship between the adopted learning strategy and student's responses to pre- and post-course survey questions regarding their prior knowledge, motivation, goal orientation, self-efficacy, and the use of specific study strategies.
- Our results revealed that there are significant differences between the identified learning strategies regarding students' levels of cognitive presence and also regarding students' commitment to learning, enrollment goals and motivation, approaches to learning, and goal orientation.

Research output:

1. Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., de Vries, P., Hatala, M., Dawson, S., Siemens, G., and Gašević, D. (2017): *"The role of technology use on shaping student learning experience in MOOCs"*: A journal article, currently under review, which describes identified learning strategies within MOOCs and their relationship to the perceived levels of cognitive, teaching, and social presence, as well as different measures of pre- and post-course survey questions relating to student motivation, self-regulation, goal orientation and approaches to learning.

1.3.7 Overview of chapter eight: "Conclusions and future directions"

Finally, in Chapter eight we examine the impact of the present work with respect to the five research questions defined in Chapter one. We also discuss the potential directions for the future works as well as for practical application of the research presented in this thesis. We finally conclude with a short overview of the thesis and a summary of its key contributions.

2 Cognitive presence assessment model

It is, in fact, nothing short of a miracle that the modern methods of instruction have not yet entirely strangled the holy curiosity of inquiry.
— Albert Einstein, *Autobiographical Notes*

2.1 Introduction

THIS chapter introduces the conceptual model for cognitive presence assessment using learning analytics methods. The model consists of four layers which together provide an overview of the key components of the analytics systems and how are combined in a methodologically sound and coherent manner. The four layers are

1. *Educational theory layer*, which provides a connection to the theoretical foundation given by the CoI model (Garrison et al., 2001).
2. *Educational technology layer*, which specifies the key data sources used for the development of learning analytics models.
3. *Assessment framework layer*, which operationalizes key components of the assessment model.
4. *Assessment approach layer*, which provides actual implementation of analytics design that depends on the specifics of a given learning context.

The methodological foundation for the model is given by Evidence-Centered Design (ECD) framework (Mislevy et al., 2003), which is widely used for the development of different forms of educational assessment. Originally developed by Educational Testing Service, the original focus of

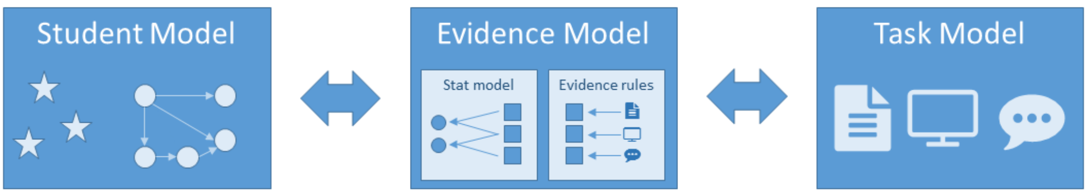


Figure 3. Conceptual assessment framework layer of evidence-centered design framework (adopted from Mislevy & Haertel, 2006).

the ECD framework was the development of large-scale standardized testing models. However, as indicated by Bond (2014), the ECD framework can be successfully used to improve the design of any form of student assessment, and as a solid foundation for formative assessment and feedback (Snow, Haertel, Fulkerson, Feng, & Nichols, 2010). The ECD framework consists of five different layers, that capture specific phases of assessment design, development, and implementation (Mislevy, Behrens, DiCerbo, & Levy, 2012). From the standpoint of this thesis, the most directly relevant part of ECD framework is the conceptual assessment framework (CAF)(Figure 3), which specifies three key elements of any assessment system. Those include:

1. *Student model*, which captures constructs measured and assessed.
2. *Task model*, which defines learning activities during which the assessment is made, and
3. *Evidence model*, which specifies the list of measures that are collected during task model activities and that provide the assessment of student model constructs.

With respect to the development of learning analytics systems, the central element is the evidence model which, as shown in Figure 3, internally consists of two distinct elements: procedures for extracting assessment measures from the learning tasks, and statistical techniques for assembling together the extracted quantitative measures. In the case of assessment models based on learning analytics, the procedures are implemented using several data mining and machine learning algorithms, while assessment measures are represented as different quantitative measures extracted from the collected educational data.

The model developed for cognitive presence assessment provides a template for the development of new learning analytics models, or for the customization of the existing models for use within new instructional contexts and scenarios. Given its grounding in the educational theory, the model can also be used to guide the development of several analytics-based models so that they provide complementary views of the target theoretical constructs. In the present thesis, the two analytical models developed for understanding cognitive presence within traditional online courses provide insights into two separate types of learning activities: 1) the ones that happen in the shared world of discourse, and 2) and the ones that happen inside the private world of student reflection (Garrison et al., 2001). In this manner, the model connects the different parts of this thesis in a coherent unit that provides a holistic view of the student cognitive presence development.

2.2 Publication: A novel model of cognitive presence assessment using automated learning analytics methods

The following section includes the verbatim copy of the following publication:

> Kovanović, V., Gašević, D., Hatala, M., and Siemens, G. (2017). *A novel model of cognitive presence assessment using automated learning analytics methods*. SRI Education. Retrieved from http://a4li.sri.com/archive/papers/Kovanovic_2017_Presence.pdf

A Novel Model of Cognitive Presence Assessment Using Automated Learning Analytics Methods

By: Vitomir Kovanović, Dragan Gašević, University of Edinburgh

Marek Hatala, Simon Fraser University

George Siemens, University of Texas at Arlington

January 2017

About Analytics for Learning (A4L)

The availability of data coming from digital learning environments is creating the possibility to measure learning like never before. The Analytics for Learning (A4L) Network is made up of researchers exploring the measurement of student learning behaviors and strategies in digital learning environments. Learn more: http://analytics4learning.org/.

IMPROVEMENT ANALYTICS

The Improvement Analytics group engages in collaborative data intensive research to help educational organizations improve learning opportunities for all students. Based within SRI Education, the Improvement Analytics group leverages broad domain expertise and diverse methodological approaches to support those working with learners of all types turn promising ideas into improvements. Across multiple partnerships, we have helped educational organizations find actionable insights within complex data sets and develop as well as productively adapt instructional innovations. Learn more: http://improvement-analytics.org

SRI Education™

A DIVISION OF SRI INTERNATIONAL

SRI Education, a division of SRI International, is tackling the most complex issues in education to identify trends, understand outcomes, and guide policy and practice. We work with federal and state agencies, school districts, foundations, nonprofit organizations, and businesses to provide research-based solutions to challenges posed by rapid social, technological and economic change. SRI International is a nonprofit research institute whose innovations have created new industries, extraordinary marketplace value, and lasting benefits to society. Learn more: http://www.sri.com/education.

This material is based upon work supported by the National Science Foundation through grant SMA-1338487. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

A Novel Model of Cognitive Presence Assessment Using Automated Learning Analytics Methods

By: Vitomir Kovanović, Dragan Gašević, University of Edinburgh
Marek Hatala, Simon Fraser University
George Siemens, University of Texas at Arlington

January 2017

Contents

Abstract

In online learning, the most widely used model which outlines students' learning experience is the community of inquiry (CoI) model. Central to the CoI model is the construct of cognitive presence, which focuses on students' development of critical and deep thinking skills and is essential for the attainment of learning outcomes.

Given the latent nature of cognitive presence, there are significant challenges related to its assessment, which currently requires manual coding of online discussion transcripts or reliance on self-reported measures using survey instruments. In this paper, we present a novel model for assessing students' development of cognitive presence using automated learning analytics techniques. Building on the foundations of evidence-centered design, we developed a flexible model for assessing students' cognitive presence based on educational trace data that can be used in variety of learning contexts (e.g., traditional for-credit online courses, massive open online courses, and blended courses). We used the model to develop two analytics systems for real-time monitoring of cognitive presence development and for delivering valuable feedback for instructors, enabling them to use different instructional interventions during a course.

Introduction

Over the course of history, technology has been redefining almost all aspects of human life (Siemens, 2005). The development of communication technologies and the availability of large amounts of digital information have brought important changes to education. Modern education is becoming extensively reliant on digital technologies, with learning management systems (LMS) redefining on-campus learning in the last 15–20 years. Similarly, more and more K–12 institutions and corporations have adopted novel technologies as a way of enhancing the learning and training experience.

Whereas technology has dramatically changed on-campus course delivery, it has created a true revolution in online and distance education, which is becoming an increasingly important mode of education delivery (Vaughan, Cleveland-Innes, & Garrison, 2013). Recent reports (GSV Advisor, 2012) have shown that in the Fall 2010 term, 6.1 million US-based higher education students were enrolled in at least one online course, and those numbers have only been rising since then. The development of massive open online courses (MOOCs)—available free to millions of students—was a global phenomenon in education and attracted significant attention from business (Friedman, 2012), academia (Gašević, Kovanović, Joksimović, & Siemens 2014), and the general public (Kovanović, Joksimović, Gašević, Siemens, & Hatala, 2015d).

Along with the introduction of modern educational technologies has been an increased interest in the development of critical and deep thinking skills. Although critical thinking has long been considered as the primary goal of education by some (e.g., Dewey, 1910), it has been attracting more attention only since the 1980s. In more recent times, critical thinking has been recognized as one of the core 21st-century skills—alongside creativity, collaboration, and communication—necessary to work in the global economy (Trilling & Fadel, 2009).

One of the widely used approaches for the development of critical thinking skills is inquiry-based learning. Rather than presenting already established facts and information in a smooth path, inquiry-based learning begins with a question, problem, or scenario. Knowledge is built through the interaction between students and the learning materials as well as with other students and instructors. It is a foundational pedagogical tool behind social-constructivist approaches to learning (Anderson & Dron, 2010) which focus on knowledge (co)-construction between learners.

In the context of online learning, one pedagogical framework that focuses on the development of critical thinking skills is the Community of Inquiry (CoI) model (Garrison, Anderson, & Archer, 1999). This model outlines three main constructs—known as *presences*—that shape students' overall online learning in communities of inquiry. *Cognitive presence* is the central construct of the CoI model and represents the operationalization of critical thinking development within an inquiry-based learning context (Garrison, Anderson, & Archer, 2001). Although cognitive presence has been recognized as important in student learning outcomes, assessing it is challenging, primarily because of its latent nature (Akyol & Garrison, 2011b). Also, the physical separation between course participants in online learning adds a layer of separation between observed behavior and the constructs of interest (Kearns, 2012). One potential approach for objective measuring of student learning and cognitive presence is through the use of fine-grained data collected by the educational technology (Shaffer et al., 2009). The field of learning analytics focuses on the utilization of this rich source of data for improvement of student learning outcomes and the learning experience (Baker & Siemens, 2013).

In this paper, we present a novel assessment model of students' cognitive presence based on the automated methods and techniques of learning analytics. The assessment model was built using the evidence-centered design framework (Mislevy, Almond, & Lukas, 2003) and was designed to collect, measure, and evaluate levels of student cognitive presence. We developed the model using several learning analytics methods, as explained in detail here.

Community of Inquiry Model

Overview

Communities of inquiry is a theoretical model that describes different dimensions that form educational experience in online learning and defines a technique for assessing the quality of the educational experience. It is based on social-constructivist ideas and is best suited for learning in higher education (Garrison et al., 1999). The CoI model attracted much attention in the research community, resulting in a significant number of replication studies (Garrison, Anderson, & Archer, 2010).The model consists of three interdependent constructs (Figure 1) that together provide comprehensive coverage of distance learning phenomena:

- *Social presence* describes the relationships and social climate in a learning community that have a significant impact on success and quality of social learning (Rourke, Anderson, Garrison, & Archer, 1999).
- *Cognitive presence* describes the different phases of students' cognitive engagement and the process of knowledge construction and development of deep thinking (Garrison et al., 2001).
- *Teaching presence* explains the instructional role during the process of social learning (Anderson, Rourke, Garrison, & Archer, 2001).

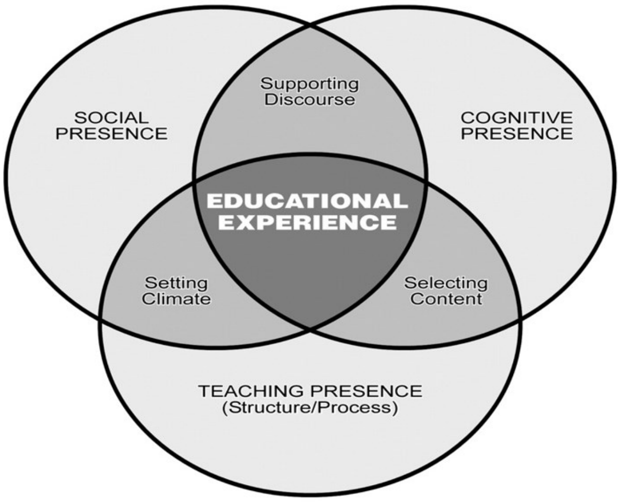


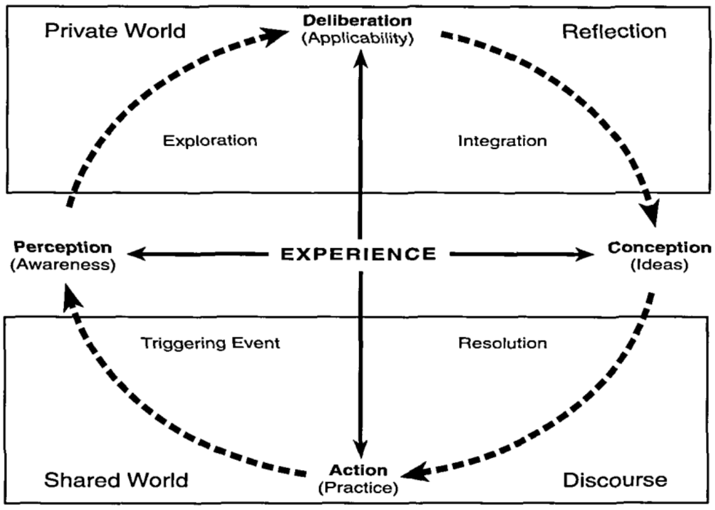
Figure 1. Community of inquiry (CoI) model

With its broad adoption in online learning, the CoI model has been used in a variety of settings that do not necessarily fall under the category of inquiry based (Garrison et al., 2010). Because the CoI model defines critical dimensions of the online learning experience, it can be used to understand and research a range of settings, including blended, lifelong, and workplace learning (Garrison et al., 2010). Likewise, the model was used as a framework for the evaluation of the different pedagogical approaches in distance and online education (Anderson & Dron, 2010).

Cognitive Presence

Cognitive presence, as defined by Garrison et al. (1999), is the "extent to which the participants in any particular configuration of a community of inquiry are able to construct meaning through sustained communication" (p. 89). It is grounded in the critical-thinking literature, most notably in the works of John Dewey and his idea that education has two sides, psychological and social, and that education is, in essence, a collaborative reconstruction of experience (Garrison et al., 1999). Cognitive presence in the CoI model is operationalized by the practical inquiry model (Garrison et al., 2001).

Figure 2. Practical inquiry model of cognitive presence



In the model, there is a clear distinction between the private (individual) world of reflective thinking and the shared (social) world of discourse and discussion, both of which are required for the development of critical thinking skills. Through dimensions of *action-deliberation* and *perception-conception*, the model defines four phases of cognitive presence:

- **Triggering event** in which an issue, dilemma, or problem is identified. In formal education, such triggers are often explicitly defined by the instructors, but any student who participates in the discussion can also do so (Garrison et al., 2001). This phase is called triggering event as it "triggers" the cycle of learning when a problem or issue regarding the practical application of knowledge is identified, which leads to the dilemma and awareness of that problem. For the quality of the learning process, the instructor has a major role in initiating or modifying triggering events and sometimes even discarding distracting ones to lead students to the desired learning outcomes. The discourse is occurring in the shared world and is the main way that students develop awareness of the dilemma or problem.
- **Exploration**, in which students are constantly moving between the private world of critical reflection and the shared world of discourse (Garrison et al., 2001). At the start of the phase, students should understand the basic nature of the problem and then move to a fuller exploration of the relevant information. As the students gain more knowledgeable at the end of this phase,

they also become more selective about the ideas that are relevant, which leads them to synthesis and integration of the explored ideas.

- **Integration**, characterized by the synthesis of the ideas that were generated in the exploration phase and ultimately by the construction of meaning (Garrison et al., 2001). This phase is, unfortunately, the hardest one to detect from the content of discourse transcripts. In general, students will tend to stay in the more “comfortable” phase of exploration, so a strong teacher presence is essential to guide critical thinking into more advanced stages through probing questions, comments, and diagnosis of misconceptions (Garrison et al., 2001).
- **Resolution**, in which the dilemma or problem is resolved by the means of direct or vicarious action (Garrison et al., 2001). In general, this is accomplished through hypothesis testing and implementation of the proposed solution. In the typical learning context, however, resolution usually involves vicarious testing through thought experiments or through building a consensus in the community as practical implementation is not feasible. The outcome of this phase is the knowledge that students are assumed to have acquired that will enable them to move to the next cycle with the triggering of the new issue, dilemma, or problem.

Overall, the cognitive presence model allows for assessment of the development of critical thinking over time within the group dynamics of communities of inquiry. The focus is on the evaluation of the *process*, and not the *product* of critical thinking, as this is much more important from the standpoint of cognitive development within communities of inquiry (Garrison et al., 2001). Instead of looking at the correctness, logic, depth, and other aspects of the products of critical thinking, the quality of the process is judged through participants’ interactions. As summarized by Garrison et al. (2001), the essential characteristic of communities of inquiry is that “members question one another, demand reasons for beliefs, and point out consequences of each other’s ideas—thus creating a self-judging community when adequate levels of social, cognitive, and teacher presence are evident” (p. 12).

Assessment of Cognitive Presence

Quantitative Content Analysis Approach

In the CoI model, the primary approach for assessing the three presences is qualitative content analysis, which is *"*a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use*"* (Krippendorff, 2003, p. 18). Qualitative content analysis is a well-established research technique often used in social science research. It usually involves using a particular coding scheme for annotation of a vast number of text documents. Before starting content analysis, researchers first define *unit of analysis* (e.g., message, paragraph, sentence, clause), which is the smallest unit of text that can be annotated, and then a code is manually assigned to each unit (De Wever, Schellens, Valcke, & Van Keer, 2006; Fahy, 2001; Strijbos, Martens, Prins, & Jochems, 2006).

For assessing the levels of cognitive presence, the CoI model has a defined coding scheme with a list of descriptors and indicators of the four phases of cognitive presence (Appendix B). Also, a list of associated socio-cognitive processes for each indicator provides a broader description of the indicators. The coding scheme has been used in a significant number of studies (c.f. Garrison et al., 2010), and in most cases sufficiently high interrater agreement was obtained (i.e., Cohen's Kappa > 0.7 which represents nearly perfect agreement).

The coding scheme was successfully used in many studies of the process of development of critical thinking in distance education. However, there are many aspects of critical thinking in general, and cognitive presence in particular, that are not addressed by this coding scheme.

Survey Instrument Approach

Besides the coding instrument, the CoI self-reported instrument for measuring the levels of the three presences is also available (Arbaugh et al., 2008). The instrument enables instructors to gather data that reflect the different presences within a course. The survey consists of 34 items on a 5-point Likert scale, with items 1–13 measuring teaching presence, items 14–22 social presence, and items 23–34 cognitive presence (Appendix C).

Evidence-Centered Design and Learning Analytics

Overview of Educational Assessment

Educational assessment deals with the measurement of student learning. Gipps (1994) defined educational assessment as "a wide range of methods for evaluating pupil performance and attainment including formal testing and examinations, practical and oral assessment, classroom-based assessment carried out by teachers and portfolios" (p. vii). Assessment represents one of the central components of education and a well-established domain of educational research and practice with roots in psychology and psychometrics. Distinctions exist between different types of assessment depending on their purposes. The purpose of *formative assessment* is to provide students and instructors feedback on learning progress so as to improve the learning outcomes (Gipps, 1994). In contrast, the goal of *summative assessment* is to examine the outcomes of the learning processes and the overall level of student learning. For example, asking a student to submit an outline of a paper is a form of formative assessment, whereas assigning a final grade to the student is a form of summative assessment.

Attention to formative assessment has grown over time, with calls for more assessment *for learning* rather than assessment *of learning* for accountability and measurement purposes. The provision of timely formative feedback has been shown to be one of the best approaches for improving student learning outcomes (Yeh, 2009). According to Shute (2008), formative feedback should be "nonevaluative, supportive, timely, and specific" (p. 153). Nicol and Macfarlane-Dick (2006) listed seven guidelines defining good formative feedback:

1. "Helps clarify what good performance is (goals, criteria, expected standards),
2. Facilitates the development of self-assessment (reflection) in learning,
3. Delivers high-quality information to students about their learning,
4. Encourages teacher and peer dialogue around learning,
5. Encourages positive motivational beliefs and self-esteem,
6. Provides opportunities to close the gap between current and desired performance, and
7. Provides information to teachers that can be used to help shape teaching" (p. 205).

In the context of online and blended learning, educational technology enables development of different systems for the provision of relevant, timely formative self-, peer-, and instructor-led feedback, which in turn helps students develop their metacognitive skills and strategies (Vaughan et al., 2013). For example, quizzes, tutoring systems, practice exams, and blogging systems are often used to help students in their self-assessment of learning. Similarly, discussion boards, collaborative writing tools, and wikis are often used for peer feedback as they allow for easy asynchronous communication among the students. Finally, instructor-led feedback—aside from rubrics and summative feedback in the form of midterm and final grades—is supported through collaborative writing tools and video conferencing, which enable instructors to provide students with comments on the quality of their learning products (Vaughan et al., 2013).

Evidence-Centered Design (ECD) Framework

Several models that outline the dimensions and elements of successful assessment have been developed (Kane & Bejar, 2014), with evidence-centered design (ECD) (Mislevy et al., 2003) being one of the more prominent and more researched. ECD is “an approach to constructing educational assessments through an evidentiary arguments” (Mislevy et al., 2003, p. i). Developed by Educational Testing Service, it is a conceptual framework designed to enhance the design and implementation of various forms of assessment, in particular, large-scale assessment (Snow, Haertel, Fulkerson, Feng, & Nichols, 2010). ECD is designed to be flexible enough to support a variety of assessment types, to be useful as a form of development checklist, and to be a tool for improving the overall quality of student assessment (Bond, 2014). According to Mislevy et al. (2003), the three fundamental premises behind ECD are:

1. “An assessment must build around the important knowledge in the domain of interest and an understanding of how that knowledge is acquired and put to use.
2. The chain of reasoning from what participants say and do in assessments to inferences about what they know, can do, or should do next, must be based on the principles of evidentiary reasoning.
3. The purpose must be the driving force behind design decisions, which reflect constraints, resources, and conditions of use” (p. 20).

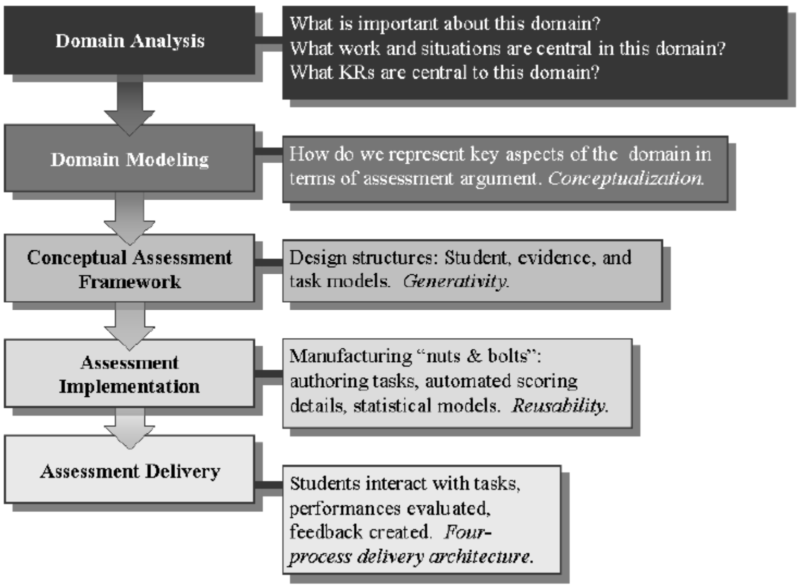
Although ECD was primarily developed for standardized test development, its generalizability and flexibility make ECD useful for a variety of assessment types, such as assessment of argumentative reading and writing skills (Deane & Song, 2014), and the solid foundation for formative feedback (Snow et al., 2010).

The ECD framework consists of five layers (Figure 3) (Mislevy, Behrens, DiCerbo, & Levy, 2012). The focus of *domain analysis* is building the understanding of a domain necessary to identify relevant constructs (e.g., knowledge, skills, tasks, activities, scenarios). *Domain modeling* builds on the identified constructs, resulting in a developed structure and identification of dependencies among them. Tangible products of domain modeling often are one or more *design patterns*, which provide the basis for specifying tangible elements of assessment design (Kane & Bejar, 2014). The third layer, *conceptual assessment framework* (CAF), outlines the critical operational elements (models) of the assessment that together coherently implement the goals of the assessment. The three core elements of the CAF framework are the

1. *Student model* (i.e., “what we measure”), which defines variables that are the target of the assessment. In our study, these were student cognitive presence and metacognition.
2. *Evidence model* (i.e., “how we measure”), which defines how we should measure the variables specified in the student model. It has two parts:
    a. *The evaluation component* provides definitions and specifications for identification and evaluation of observed variables (Mislevy et al., 2003).
    b. *Measurement model*, which provides a link between variables in student model and observed variables and their values. In the simplest form, measures of observed variables can be plain summed scores, but more complex models using Bayesian statistics or item-response theory can be defined (Mislevy et al., 2012).
3. *Task model* (i.e., “when we measure”), defines the structure of activities so that the evidence of student performance related to variables of interest can be adequately acquired (Mislevy et al., 2003).

The last two layers concern practical issues of assessment implementation and delivery. The *assessment implementation layer* focuses on the authoring of assessment tasks (or development of automated systems for their production), the specification of scoring, statistical models, model testing, estimation of model parameters, and other implementation details. Finally, *assessment delivery* defines the *four-process delivery architecture* for the actual assessment orchestration and its practical use.

Figure 3. Evidence-centered design framework (Mislevy et al., 2012)



Assessment Through Automated Learning Analytics

According to Siemens, Long, Gašević, and Conole (2011), learning analytics is “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (para. 4). It is a multidisciplinary research area that draws heavily from machine learning and data mining (Cooper, 2012) with the intention to build systems and tools that can “inform and empower instructors and learners, such as informing instructors about ways that specific students are struggling, so that the instructor can contact the learner” (Baker & Siemens, 2013, p. 4).

Although learning analytics is still a very young research field, its potential to impact educational practice and research through student assessment has been recognized (Ellis, 2013). As summarized by Ellis (2013), “Assessment analytics offers the potential for student attainment to be measured across time, in comparison with individual students’ starting points (ipsative development), with their peers and/or against benchmarks or standards” (p. 663). Indeed, over time a large number of studies have investigated the use of automated analytical systems for assessment of student learning, including assessment of student essays (Dikli, 2006; Duwairi, 2006; Foltz et al., 1999), reading comprehension (Allen, Snow, & McNamara, 2015; Mintz, Stefanescu, Feng, D’Mello, & Graesser, 2014), digital literacies (Dawson & Siemens, 2014), programming skills (Blikstein, 2011), self-assigned grades (Fuentes, Romero, & Ventura, 2014), and social capital (Joksimović, et al., 2016; Joksimović, Gašević, Kovanović, Riecke, & Hatala, 2015a). Learning analytics is also used extensively to predict students’ final course outcomes and to identify “at-risk” students (Ferguson, 2012).

In a similar manner, the generalizability of the ECD makes it possible to include automated elements in the assessment design (Mislevy et al., 2012), as already demonstrated by several studies (Behrens, Mislevy, Bauer, Williamson, & Levy, 2004; Rupp et al., 2012; Shaffer et al., 2009). Although ECD was originally developed for traditional multiple-choice questions, it can be used in more complex scenarios (Behrens et al., 2004) where automated data mining and learning analytics can provide richer data for more complex assessment that moves beyond traditional item scoring. Learning analytics and educational data mining can serve to identify variables indicative of the latent constructs of interest and improve the quality of the evidence used for student assessment (Mislevy et al., 2012). Finally, studies have indicated the utility of the ECD framework in developing models of formative assessment (Shute, 2004; Shute, Ventura, Bauer, & Zapata-Rivera, 2009; Snow et al., 2010).

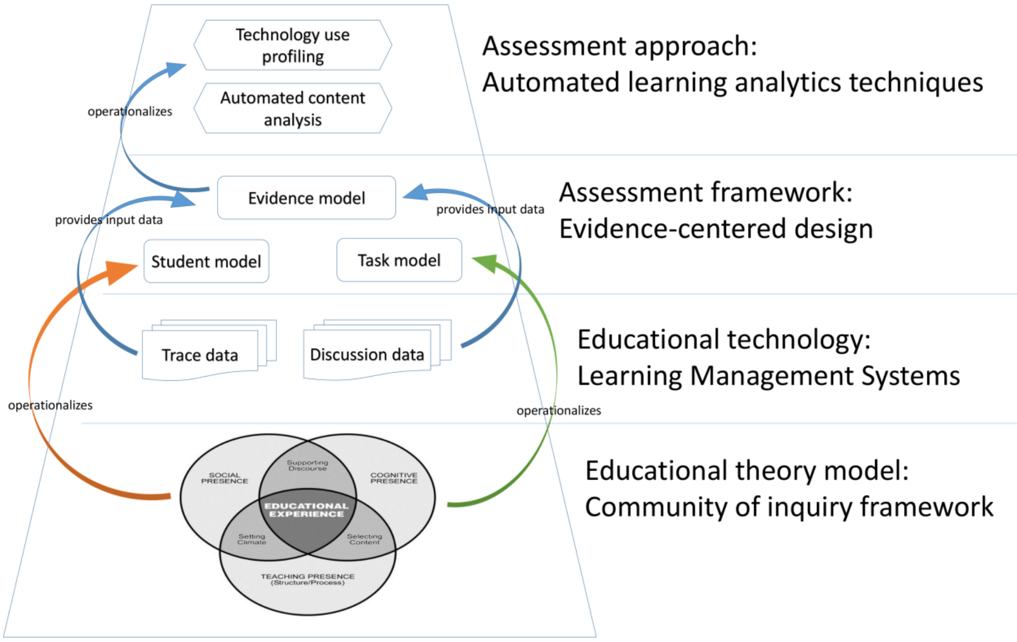
Cognitive Presence Assessment Model

Here, we describe the integral parts of the ECD-based assessment model. The model is conceptually outlined in Figure 4. The theoretical foundations of social-constructivist learning operationalized by the CoI model are at the bottom of the pyramid. Next is the technological layer, which provides input data for the particular assessment. Building on these foundations is ECD-based assessment framework layer, with student model, task model, and evidence model being the three key components. Finally, from these three models, we developed two analytical approaches, as an implementation of the described assessment model.

Student Model

The goal of student model in the ECD framework is to operationalize the knowledge, skills, or other attributes (KSAs) that are the target of the assessment (Mislevy et al., 2003). In our study, the central construct that was evaluated was student cognitive presence. Besides cognitive presence, important KSAs were students' prior domain knowledge, self-efficacy (Bandura, 1977; Shea & Bidjerano, 2010), self-regulation of learning (Butler & Winne, 1995), metacognition (Akyol & Garrison, 2011a; Azevedo & Aleven, 2013; Flavell, 1979), goal orientation (Meece, Blumenfeld, & Hoyle, 1988; Senko, Hulleman, & Harackiewicz, 2011), motivation (Ames, 1992; Deci, Vallerand, Pelletier, & Ryan, 1991; Hartnett, George, & Dron, 2011; Kizilcec & Schneider, 2015), digital literacy (Gilster, 1997), and familiarity with the available technological tools. It might be the case, for example, that a student who exhibits lower cognitive presence is facing challenges with a particular study domain or adopted learning technology. Similarly, individual differences in goal orientation and motivation will most likely be reflected in their study approaches and regulation of their learning activities (Biggs, Kember, & Leung, 2001).

Figure 4. Conceptual diagram of the framework for assessment of cognitive presence using learning analytics



Task Model

The task model defines the activities and tasks to be used to provide evidence about the constructs specified in the student model. For cognitive presence assessment—given the social-constructivist underpinning of the learning with the CoI model—there are two broad groups of activities: private-world self-reflective learning tasks and shared-world social learning tasks.

The first group consists of activities that are indicative of students' individual learning. Those include accessing course resources, taking practice exams, watching lecture recordings, and producing essays, video presentations, wiki pages, blog posts, and other types of text/video content. The list of activities in the first group will depend on the design and organization of a particular course (Gašević, Dawson, Rogers, & Gasevic, 2016). For example, in a traditional online course, it is very unlikely that students would write blog posts, whereas in a connectivist MOOC (cMOOC) that would be a very common activity (Joksimović et al., 2015b). The particular course design choices will have an impact on the design elements that will be included in the evidence model and subsequently provide evidence of student learning.

The second group of activities consists of students' discourse within online discussion forums. Those involve reading other students' messages and posting new messages and message replies. Given that the use of online discussions is essential for the social-constructivist pedagogies and the foundation of inquiry-based learning, online discussions and their use are the primary targets of the current content analysis approaches. The course design also plays a major role in creating rules and setting up students' expectations of their participation (Gašević, Adesope, Joksimović, & Kovanović, 2015), as the mere provision of the technological affordances for online discussions in most cases is not sufficient.

Evidence Model

The evidence model provides instructions on how to gather the information about the variables described in student model from the execution of the tasks and activities defined in the task model (Mislevy et al., 2003). The *evaluation component* (also called *evaluation rules*) of the evidence model defines how identification and evaluation of the observed variables should be conducted, whereas the *measurement model* specifies the connection between student model variables and the observed variables (Mislevy et al., 2003).

In our context, the evaluation component consists of the list of observed variables extracted from the LMS trace data and online discussion data. From the LMS trace data, the primary observed variables are individual event records of student actions defined in the task model. Those include trace records of student course logins, discussion views, viewing of course resources, quiz attempts, and other events prescribed by the course design. From discussion data, the primary variables that are indicative of student model variables are discussion post contents and discussion post metadata (i.e., date and time of posting, discussion topic name, the list of previous topic messages). The evaluation component simply accumulates the list of events for a particular student, which are then used in the measurement model to define appropriate measures of student model variables.

Based on the evidence rules, the measurement model for trace data consists of two types of measures: (1) *count measures*, which provide an indication of how many times a particular action occurred for a given student, and (2) *time-on-task measures* (Kovanović et al. 2015a, 2015b), which indicate how much *time* a student spent on a particular type of activity. Count measures included variables such as the number of logins, the number of course page views, the number of resource downloads/views, the number of discussions, and other measures related to different parts of the LMS. Most of the extracted count measures have corresponding time-on-task measures (e.g., time spent viewing course pages, time spent viewing resources). As indicated by Kovanović et al. (2015a, 2015b), there are a small number of "instantaneous" measures that do not have a meaningful corresponding time-on-task measure (e.g., logging into the LMS system, running a search on the discussion data, marking a discussion as read, subscribing to the discussion updates). From online discussion data, the measurement model consisted of the different text classification features, which were extracted from the list of student online postings and its metadata. Those included (1) measures of message context and position within threaded discussion (e.g., message ordinal number in the thread, 0-1 indicator [whether the message was the first or last in the thread], similarity with the previous message), (2) large number of different linguistic measures (i.e., text cohesion, count of words in the various psychological categories), and (3) message content features (e.g., length, number of content concepts).

Empirical Validation of the Framework

The model was used in several studies to develop two different learning analytics assessments of student learning within the community of inquiry model. The study by Kovanović, Gašević, Joksimović, Hatala, and Adesope (2015c) built on the proposed model to define a student-clustering model that provided insights into students' use of the available LMS tools as an indicator of their learning regulation. The student model consisted of student cognitive presence, and the task model consisted of (1) viewing and posting to online discussions, (2) using online quizzes, (3) submitting assessments, and (4) using online course resources. The evaluation model consisted of thirteen variables from the two groups of activities, private-world and shared-world activities:

A. Private-world self-learning activities
    1. `UserLoginCount`: the number of times student logged into the system.
    2. `CourseViewCount`: the number of times student opened course information pages
    3. `AssignmentViewTime`: the time spent on course assignments
    4. `AssignmentViewCount`: the number of times student opened assignment pages
    5. `ResourceViewTime`: the time spent reading online resources
    6. `ResourceViewCount`: the number of times student opened one of the course resources
B. Shared-world discussion-related measures
    7. `ForumSearchCount`: the number of times student searched in online discussions
    8. `DiscussionViewTime`: the time spent viewing online discussions
    9. `DiscussionViewCount`: the number of times student opened online discussions
    10. `AddPostTime`: the time spent posting discussion messages
    11. `AddPostCount`: the number of discussion board messages posted by the student
    12. `UpdatePostTime`: the time spent updating discussion messages
    13. `UpdatePostCount`: the number of times student updated one of his or her discussion messages.

Using the defined student, task, and evaluation model, Kovanović et al. (2015c) developed an automated clustering system that can be used to detect study strategies indicative of student cognitive presence development. The study identified six different study strategies that differed in levels of cognitive presence development, with studies that included an online discussion component showing higher cognitive presence development than studies that focused primarily on individual learning activities.

Another study in which the proposed conceptual framework was used was the one by Kovanović et al. (2016) on the social component of the cognitive presence development. In that case, the task model was only online discussion posting and viewing. The evaluation portion of the evidence model consisted of discussion message contents and associated metadata, whereas the measurement model consisted of 205 measures extracted from the discussion message content and metadata. Those measures included

- 108 LIWC (Linguistic Inquiry and Word Count) features (Tausczik & Pennebaker, 2010), which is a set of word counts in different linguistic categories (e.g., positive/negative emotional words, cognitive words, pronouns, social words, perceptual words)
- 205 Coh-Metrix (McNamara, Graesser, McCarthy, & Cai, 2014), which is a set of measures related to the cohesion of the written text.
- Six discussion context features—number of replies, message depth (i.e., thread ordinal position), cosine similarity with previous/next message, indicator of first/last message in the discussion thread
- Message content features—number of named entities extracted using DBPedia Spotlight (Mendes, Jakob, García-Silva, & Bizer, 2011), message length, and average Latent Semantic Analysis (LSA) similarity of message paragraphs (i.e., how similar paragraphs of a message are).

Using the set of measures, Kovanović et al. (2016) developed a learning analytics system that can automatically detect the level of cognitive presence in each discussion message. Through automated text mining techniques, Kovanović and colleagues developed a system that classifies each message to one of the four phases of cognitive presence, which is then used to assess the student’s development of cognitive presence.

Summary and Contributions

In this paper, we presented a novel model for assessing levels of cognitive presence in communities of inquiry based on automated learning analytics techniques. Using evidence-centered design as the theoretical foundation, we developed an assessment model and a set of automated learning analytics tools that can be used to provide rich and holistic forms of assessment of students' cognitive presence development. The flexibility of the assessment model and the automated nature of the analytics tools are significant improvements over current approaches for cognitive presence assessment, and this study contributed to advancements in research on and practice with the CoI model.

Although the development of critical thinking involves both individual learning (i.e., private-world learning) and social learning (i.e., shared-world learning), the current models of assessment based on content analysis look only at cognitive presence development as expressed in transcripts of online discussions. As students' use of online learning systems involves more than just the use of online discussions, examining the LMS trace data records can provide insights into individual learning activities and learning self-regulation, which can be then used to explain the observed levels of critical thinking in the discussion transcripts.

The use of automated analytics techniques for assessment enables continuous monitoring of cognitive presence development, which instructors can use to alter their instructional approaches during a course and in turn improve student learning outcomes. Current content analysis and self-reported instruments do not allow for this type of feedback, primarily because of their high costs and invasiveness, respectively. Automation of cognitive presence assessment also opens the door for more personalized learning experiences and individually tailored instructional interventions. For example, a student's cognitive presence can be monitored with regard to different course topics or learning objectives, which can give instructors cues for what parts of the curriculum students may require additional instructional support on. At present, this is not commonly done as the existing assessment instruments are almost exclusively administered post-course and examine cognitive presence at the whole-course level.

The use of learning analytics for the assessment of cognitive presence eases adoption of the CoI model by practitioners and researchers and in a wider set of learning contexts. The existing content analysis methods are very time-consuming, expensive, and require—aside from knowledge of the CoI model—special training in the CoI coding scheme before acceptable levels of interrater agreement are reached. The use of automated methods allows for much simpler, easier, and richer monitoring of student cognitive presence development, which improves the potential adoption of CoI model by the researchers and practitioners. Automation is particularly important for settings such as MOOCs, where the particularly large number of students makes it very hard to assess cognitive presence using existing instruments. Finally, by being automated and based on tracked evidence of student activities, learning analytics assessment models provide more objective validation of student learning, unlike transcript coding or self-reporting.

From a theoretical perspective, the developed assessment models provide further insights into the CoI model. Given the data-driven nature of developed assessment models, they provide evidence-based operationalization of the CoI constructs with data available in discussion transcripts and other types of digital traces such as clickstream data in LMSs. As the CoI model and its instruments provide very high-level conceptual descriptions of the phases of cognitive presence, automated models can be used to provide more precise data-driven operationalization of the cognitive presence construct. For example, how is a sense of puzzlement (indicative of triggering even phase) shown in discussion transcripts or trace data? Similarly, how is divergence (in a message or community), which is indicative of the exploration phase, expressed on the linguistic level? These and similar questions are implicitly answered by developing automated data-driven learning analytics assessment techniques.

Developing assessment models for constructs such as cognitive presence is a significant step toward more comprehensive models for student assessment. For a long time, there have been calls for shifting the focus of assessment from final grades and item-based testing to assessment for learning. This is especially important given the recent developments in online education, MOOCs, and the overall rise of not-for-credit learning, where there are no final grades for learners and no summative assessment in the traditional sense. Nonetheless, it is still important to provide instructors and students with (formative and summative) feedback that would improve both learning outcomes and learning experience. By means of learning analytics and assessment of cognitive presence, we made one step toward this important goal.

References

Akyol, Z., & Garrison, D. R. (2011a). Assessing metacognition in an online community of inquiry. *The Internet and Higher Education*, *14*(3), 183–190. doi:10.1016/j.iheduc.2011.01.005

Akyol, Z., & Garrison, D. R. (2011b). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology*, *42*(2), 233–250. doi:10.1111/j.1467-8535.2009.01029.x

Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Are you reading my mind?: Modeling students' reading comprehension skills with natural language processing techniques. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 246–254). New York, NY: ACM. doi:10.1145/2723576.2723617

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, *84*(3), 261–271. doi:10.1037/0022-0663.84.3.261

Anderson, T., & Dron, J. (2010). Three generations of distance education pedagogy. *International Review of Research in Open and Distance Learning*, *12*(3), 80–97. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/890/1663

Anderson, T., Rourke, L., Garrison, D. R., & Archer, W. (2001). Assessing teaching presence in a computer conferencing context. *Journal of Asynchronous Learning Networks*, *5*, 1–17.

Arbaugh, J. B., Cleveland-Innes, M., Diaz, S. R., Garrison, D. R., Ice, P., Richardson, J. C., & Swan, K. P. (2008). Developing a community of inquiry instrument: Testing a measure of the community of inquiry framework using a multi-institutional sample. *The Internet and Higher Education*, *11*(3–4), 133–136. doi:10.1016/j.iheduc.2008.06.003

Azevedo, R., & Aleven, V. (2013). Metacognition and learning technologies: An overview of current interdisciplinary research. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 1–16). New York, NY: Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-5546-3_1

Baker, R., & Siemens, G. (2013). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd Ed., pp. 253–274). Cambridge, England: Cambridge University Press.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. doi:10.1037/0033-295X.84.2.191

Behrens, J. T., Mislevy, R. J., Bauer, M., Williamson, D. M., & Levy, R. (2004). Introduction to evidence centered design and lessons learned from its application in a global e-learning program. *International Journal of Testing*, *4*(4), 295–301. doi:10.1207/s15327574ijt0404_1

Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, *71*(1), 133–149. doi:10.1348/000709901158433

Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 110–116). New York, NY: ACM. doi:10.1145/2090116.2090132

Bond, L. (2014). A brief note on evidence-centered design as a mechanism for assessment development and evaluation. *Measurement: Interdisciplinary Research and Perspectives*, *12*(1–2), 37–38. doi:10.1080/15366367.2014.921486

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245–281. doi:10.3102/00346543065003245

Cooper, A. (2012). A brief history of analytics. *CETIS Analytics Series*, *1*(9). London, UK: JISC. Retrieved from http://publications.cetis.org.uk/wp-content/uploads/2012/12/Analytics-Brief-History-Vol-1-No9.pdf

Dawson, S., & Siemens, G. (2014). Analytics to literacies: The development of a learning analytics framework for multiliteracies assessment. *International Review of Research in Open and Distance Learning*, *15*(4). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1878

Deane, P., & Song, Y. (2014). A case study in principled assessment design: Designing assessments to measure and support the development of argumentative reading and writing skills. *Psicología Educativa*, *20*(2), 99–108. doi:10.1016/j.pse.2014.10.001

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*, *46*(1), 6–28. doi:10.1016/j.compedu.2005.04.005

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist*, *26*(3–4), 325–346. 10.1080/00461520.1991.9653137

Dewey, J. (1910). *How we think*. Boston, MA: D.C. Heath & Company.

Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning and Assessment*, *5*(1). Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1640

Duwairi, R. M. (2006). A framework for the computerized assessment of university student essays. *Computers in Human Behavior*, *22*(3), 381–388. doi:10.1016/j.chb.2004.09.006

Ellis, C. (2013). Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics. *British Journal of Educational Technology*, *44*(4), 662–664. doi:10.1111/bjet.12028

Fahy, P. J. (2001). Addressing some common problems in transcript analysis. *International Review of Research in Open and Distance Learning*, *1*(2). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/321

Ferguson, R. (2012). *The state of learning analytics in 2012: A review and future challenges* (No. KMI-2012-01). Knowledge Media Institute, Open University, UK.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. doi:10.1037/0003-066X.34.10.906

Foltz, P. W., Laham, D., Landauer, T. K., Foltz, P. W., Laham, D., & Landauer, T. K. (1999). Automated essay scoring: applications to educational technology (Vol. 1999, pp. 939–944). Presented at the EdMedia: World Conference on Educational Media and Technology. Retrieved from http://www-psych.nmsu.edu/~pfoltz/reprints/Edmedia99.html

Friedman, T. L. (2012, May 15). Come the revolution. *New York Times*. Retrieved from http://www.nytimes.com/2012/05/16/opinion/friedman-come-the-revolution.html

Fuentes, J., Romero, C., & Ventura, C. G.-M. (2014). Accepting or rejecting students' self-grading in their final marks by using data mining. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014)* (pp. 327–328). International Educational Data Mining Society. Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/posters/3_EDM-2014-Poster.pdf

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer Conferencing in higher education. *The Internet and Higher Education*, *2*(2–3), 87–105. doi:10.1016/S1096-7516(00)00016-6

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23. doi:10.1080/08923640109527071

Garrison, D. R., Anderson, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education*, *13*(1–2), 5–9. doi:10.1016/j.iheduc.2009.10.003

Gašević, D., Adesope, O., Joksimović, S., & Kovanović, V. (2015). Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, *24*, 53–65. doi:10.1016/j.iheduc.2014.09.006

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. doi:10.1016/j.iheduc.2015.10.002

Gašević, D., Kovanović, V., Joksimović, S., & Siemens, G. (2014). Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative. *International Review of Research in Open and Distributed Learning*, *15*(5). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1954

Gilster, Paul. (1997). *Digital literacy*. New York, NY: John Wiley & Sons.

Gipps, C. V. (1994). Beyond testing: Towards a theory of educational assessment. London, England: Routledge.

GSV Advisor. (2012). *Education Sector Factbook 2012*. Chicago, IL: GSV Advisor. Retrieved from http://gsvadvisors.com/wordpress/wp-content/uploads/2012/04/GSV-EDU-Factbook-Apr-13-2012.pdf

Hartnett, M., George, A. S., & Dron, J. (2011). Examining motivation in online distance learning environments: Complex, multifaceted and situation-dependent. *The International Review of Research in Open and Distance Learning*, *12*(6), 20–38. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1030/1954

Joksimović, S., Gašević, D., Kovanović, V., Riecke, B. E., & Hatala, M. (2015a). Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning*, 31: 638–654. doi:10.1111/jcal.12107

Joksimović, S., Kovanović, V., Jovanović, J., Zouaq, A., Gašević, D., & Hatala, M. (2015b). What do cMOOC participants talk about in social media?: A topic analysis of discourse in a cMOOC. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 156–165). New York, NY: ACM. doi:10.1145/2723576.2723609

Joksimović, S., Manataki, A., Gašević, D., Dawson, S., Kovanović, V., & Kereki, I. F. de. (2016). Translating network position into performance: Importance of centrality in different network configurations. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 314–323). New York, NY: ACM. doi:10.1145/2883851.2883928

Kane, M. T., & Bejar, I. I. (2014). Cognitive frameworks for assessment, teaching, and learning: A validity perspective. *Psicología Educativa*, *20*(2), 117–123. doi:10.1016/j.pse.2014.11.006

Kearns, L. R. (2012). Student assessment in online learning: Challenges and effective practices. *Journal of Online Learning and Teaching*, *8*(3), 198–208. Retrieved from http://jolt.merlot.org/vol8no3/kearns_0912.htm

Kizilcec, R. F., & Schneider, E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Transactions on Computer-Human Interaction*, *22*(2). doi:10.1145/2699735

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R., & Hatala, M. (2015a). Does time-on-task estimation matter? Implications on validity of learning analytics findings. *Journal of Learning Analytics, 2*(3), 81–110. doi:10.18608/jla.2015.23.6

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015b). Penetrating the black box of time-on-task estimation. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 184–193). New York, NY: ACM. doi:10.1145/2723576.2723623

Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015c). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, *27*, 74–89. doi:10.1016/j.iheduc.2015.06.002

Kovanović, V., Joksimović, S., Gašević, D., Siemens, G., & Hatala, M. (2015d). What public media reveals about MOOCs: A systematic analysis of news reports. *British Journal of Educational Technology*, *46*(3), 510–527. doi:10.1111/bjet.12277

Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 15–24). New York, NY: ACM. doi:10.1145/2883851.2883950

Krippendorff, K. H. (2003). *Content analysis: An introduction to its methodology* (2nd ed.). New York, NY: Sage Publications.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, England: Cambridge University Press.

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, *80*(4), 514–523. doi:10.1037/0022-0663.80.4.514

Mendes, P. N., Jakob, M., García-Silva, A., & Bizer, C. (2011). DBpedia spotlight: Shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems* (pp. 1–8). New York, NY: ACM. doi:10.1145/2063518.2063519

Mintz, L., Stefanescu, D., Feng, S., D'Mello, S., & Graesser, A. (2014). Automatic assessment of student reading comprehension from short summaries. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014)* (pp. 333–334). International Educational

Data Mining Society. Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/posters/9_EDM-2014-Poster.pdf.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*, *2003*(1), i-29. doi:10.1002/j.2333-8504.2003.tb01908.x

Mislevy, R. J., Behrens, J. T., DiCerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining*, *4*(1), 11–48. Retrieved from http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/22

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218. doi:10.1080/03075070600572090

Poquet, O., Kovanović, V., Vries, P. de, Hennis, T., Joksimović, S., Gašević, D., & Dawson, S. (2016). Social presence in massive open online courses. Manuscript submitted for publication.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. *Journal of Distance Education*, *14*(2), 50–71.

Rupp, A. A., Levy, R., DiCerbo, K. E., Sweet, S. J., Crawford, A. V., Calico, T., … Behrens, J. T. (2012). Putting ECD into practice: The interplay of theory and data in evidence models within a digital learning environment. *Journal of Educational Data Mining*, *4*(1), 49–110. Retrieved from http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/23

Senko, C., Hulleman, C. S., & Harackiewicz, J. M. (2011). Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. *Educational Psychologist*, *46*(1), 26–47. doi:10.1080/00461520.2011.538646

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., … Mislevy, R. (2009). Epistemic network analysis: A Prototype for 21st-century assessment of learning. *International Journal of Learning and Media*, *1*(2), 33–53. doi:10.1162/ijlm.2009.0013

Shea, P., & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computers & Education*, *55*(4), 1721–1731. doi:10.1016/j.compedu.2010.07.017

Shute, V. J. (2004). Towards automating ECD-based diagnostic assessments. *Technology, Instruction, Cognition, and Learning*, *2*, 1–18. Retrieved from http://www.oldcitypublishing.com/FullText/TICLfulltext/TICL2.1-2fulltext/TICLv2n1-2p1-18Shute.pdf

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. doi:10.3102/0034654307313795

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). New York, NY: Routledge.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. In *International Journal of Instructional Technology and Distance Learning* *2*(1). Retrieved from http://www.itdl.org/journal/jan_05/article01.htm

Siemens, G. (2012). MOOCs are really a platform. *ELearnSpace Blog*. Retrieved from http://www.elearnspace.org/blog/2012/07/25/moocs-are-really-a-platform/

Siemens, G., Long, P., Gašević, D., & Conole, G. (2011). Call for papers, 1st International Conference Learning Analytics & Knowledge (LAK 2011). Retrieved from https://tekri.athabascau.ca/analytics/call-papers

Snow, E., Haertel, G., Fulkerson, D., Feng, M., & Nichols, P. (2010). Leveraging evidence-centered assessment design in large-scale and formative assessment practices. In *Proceedings of the 2010 Annual Meeting of the National Council on Measurement in Education (NCME)*. Retrieved from http://pact.sri.com/downloads/Leveraging-Evidence-Centered-Assessment-Design.pdf

Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education*, *46*(1), 29–48. doi:10.1016/j.compedu.2005.04.002

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. doi:10.1177/0261927X09351676

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. Hoboken, NJ: John Wiley & Sons.

Vaughan, N. D., Cleveland-Innes, M., & Garrison, D. R. (2013). Teaching in blended learning environments: Creating and sustaining communities of inquiry. Edmonton, AB: AU Press.

Yeh, S. S. (2009). The cost-effectiveness of raising teacher quality. *Educational Research Review*, *4*(3), 220–232. doi:10.1016/j.edurev.2008.06.002

Appendix A: ECD Design Pattern

Author	
First Name	Vitomir
Last Name	Kovanović
Affiliation	The University of Edinburgh
E-Mail	v.kovanovic@ed.ac.uk
Overview	
Summary	• Cognitive presence is a central construct in the Community of Inquiry (CoI) model (Garrison et al., 1999) concerning students' development of critical thinking and deep thinking skills. • Cognitive presence is specifically related to online and distance learning, especially to the traditional learning management systems (LMS)-driven for-credit online courses. However, cognitive presence can be applied more broadly to any online learning experience. • Data sources include (1) traces collected by learning management systems and include records of the different activities that students performed as well as (2) online discussions (the content of discussion messages and their metadata).
	• Community of Inquiry model is introduced by Garrison et al. (1999), while cognitive presence is operationalized by Garrison et al. (2001)
Rationale	• Cognitive presence is the key construct in the widely used community of inquiry model of online learning, and is therefore, of a direct importance for student learning through social knowledge construction. By developing cognitive presence, students develop critical thinking and deep thinking skills, which are the key graduate skills identified by many higher education institutions and are part of the larger group of so-called "21st-century skills" that are deemed essential for success in the modern global economy.
	• The primary purpose of assessing levels of cognitive presence is to provide formative feedback to both instructors and students. From the instructor's perspective, insights into students' development of cognitive presence are crucial as they guide them in modifying and altering their instructional approach. From the students' perspective, the feedback related to the development of cognitive presence could be used to provide them with actionable real-time recommendations about how to improve their study approach. The feedback is of particular importance in massive open online courses, where a large number of students makes it hard for the instructors to intervene on the individual-student level.
Student Model	

Focal Construct	• Cognitive presence, which is defined as the "extent to which the participants in any particular configuration of a community of inquiry are able to construct meaning through sustained communication" (Garrison et al., 1999, p. 89).
	• Cognitive presence is theorized to develop through four distinct phases:
	    ○ *Triggering event*—The cycle of learning is triggered by a problem, issue, or dilemma.
	    ○ *Exploration*—Students explore, brainstorm, and collect potentially relevant information on the given problem.
	    ○ *Integration*—Students synthesize the relevant information and start building solutions.
	    ○ *Resolution*—The developed solutions are applied or tested on the original problem. This phase often triggers a new learning cycle.

Additional Knowledge, Skills, and Abilities	• Prior-knowledge • Self-efficacy • Self-regulation of learning • Metacognition • Motivation • Goal orientation • Familiarity with educational technology • Digital literacy

Task Model	

Characteristic Features of the Task	• The course should be fully online or blended/hybrid.
	• Students should be using an LMS that is recording data about their activities within the system.
	    o The trace data about the different learning activities (e.g., logins, page views, online quizzes)
	    o The content of their online discussions and associated metadata (i.e., date and time of posting, author, discussion name, whether a message is a reply or not, a "source" message if the message is a reply)

Variable Features of the Task	• Variables extracted from the tools available in learning management systems, as specified by the course design:
	    o Use of online quizzes
	    o Use of video lecture recordings
	    o Use of blogs/wikis
	    o Use of course assignment submissions
	    o Use of text recourses
	    o Use of online discussions
	    o The design of online discussions
	    o The overall course grading rubric

Potential Task Products	• Student online discussions
	    o Content of all messages (i.e., message text)
	    o Context of all messages (i.e., message position within discussions, time, date, and author information)
	• Trace data recordings of the learning management system use
	    o Count measures
	    o Time-on-task measures

Evidence Model	

Potential Observations	• The total number of times each type of activity (e.g., system log-in, course view, quiz attempt, discussion view) was executed by each student. Also, the total time spent on each type of activity available in the course.
	• The content and metadata of all student discussion messages
	    o Text cohesiveness metrics (i.e., Co-Metrix variables)
	    o Number of content-related words
	    o Average paragraph similarity based on latent semantic analysis
	    o Number of words in different psychological categories (i.e., variables of the Linguistic Inquiry and Word Count framework)
	    o Discussion context features (position within the thread, similarity with previous/next message, first/last message)

Potential Frameworks	• Develop an automated learning analytics system that can detect students' levels of cognitive presence based on the data gathered by LMSs.

Appendix B: Cognitive Presence Coding Scheme

The cognitive presence coding scheme, as defined by Garrison et al. (2001).

Phase	Descriptor	Indicator	Socio-cognitive process
Triggering Event	Evocative	Recognizing the problem	Presenting background information that culminates in a question
		Sense of puzzlement	Asking questions Messages that take discussion in new direction
Exploration	Inquisitive	Divergence—within the community	Unsubstantiated contradiction of previous ideas
		Divergence—within a single message	Many different ideas/themes presented in one message
		Information exchange	Personal narratives/descriptions/facts (not used as evidence to support a conclusion)
		Suggestions for consideration	Author explicitly characterizes message as exploration—e.g., “Does that seem about right?” or “Am I way off the mark?”
		Brainstorming	Adds to established points but does not systematically defend/justify/develop addition
		Leaps to conclusions	Offers unsupported opinions
Integration	Tentative	Convergence—among group members	Reference to previous message followed by substantiated agreement, e.g., “I agree because...”
		Convergence—within a single message	Justified, developed, defensible, yet tentative hypotheses
		Connecting ideas, synthesis	Integrating information from various sources—textbook, articles, personal experience
		Creating solutions	Explicit characterization of message as a solution by participant
Resolution	Committed	Vicarious application to real world	None
		Testing solutions	Coded
		Defending solutions	

Appendix C: Cognitive Presence Survey Instrument

The survey items related to cognitive presence, as defined by Arbaugh et al., (2008) are:

A. Triggering event questions:
   1) Problems posed increased my interest in course issues.
   2) Course activities piqued my curiosity.
   3) I felt motivated to explore content related questions.

B. Exploration questions:
   1) I utilized a variety of information sources to explore problems posed in this course.
   2) Brainstorming and finding relevant information helped me resolve content related questions.
   3) Online discussions were valuable in helping me appreciate different perspectives.

C. Integration questions:
   1) Combining new information helped me answer questions raised in course activities.
   2) Learning activities helped me construct explanations/solutions.
   3) Reflection on course content and discussions helped me understand fundamental concepts in this class.

D. Resolution questions:
   1) I can describe ways to test and apply the knowledge created in this course.
   2) I have developed solutions to course problems that can be applied in practice.
   3) I can apply the knowledge created in this course to my work or other non-class- related activities.

A4L Analytics for Learning

2.3 Summary

In this chapter, we gave an overview of the conceptual model of cognitive presence assessment developed using ECD framework. The model serves as a foundation for the work presented in this thesis and shows how the different analytics fit together to provide a comprehensive overview of the cognitive presence from both personal (reflective) and social (discourse) perspective. Most importantly, the model provides detailed description of the three key elements of assessment design, which include 1) operationalization of key theoretical constructs related to cognitive presence (i.e., student model), 2) specification of the key learning activities in which assessment of cognitive presence can be realized (i.e., task model), and 3) the list of quantitative measures of students' cognitive presence which can be obtained from the given list of learning activities (i.e., evidence model).

The cognitive presence assessment model described in this chapter serves as a template for the development of new forms of cognitive presence analytics, as well as for the adoption of the existing analytical systems to new learning contexts. For instance, in Chapter three and Chapter four, we describe two particular learning analytics implementations focused on assessing cognitive presence in traditional, for-credit online courses that use asynchronous discussions as the primary means of social communication. Similarly, in Chapter seven, we present a new learning analytics model for assessing cognitive presence in MOOCs which is heavily influenced by the similar model from Chapter four, developed for the use within traditional online courses. Given the importance of course organization and design for the development of student cognitive presence, the present model can be adjusted to enable the assessment of cognitive presence for a broad range of learning contexts, such as blended or flipped classroom courses.

There are some important contributions of the work presented in this chapter – aside from its use in guiding design and development of cognitive presence assessment models. First, to the best of our knowledge, the model presented is the first model to describe assessment of cognitive presence from the standpoint of personal, reflective dimension of inquiry-based learning (Garrison et al., 2001). Secondly, the model developed for cognitive presence assessment is the first step towards the more qualitative measurement of student learning outcomes in a data-informed and objective manner. With the rising importance of non-formal educational contexts, such as MOOCs, the move away from simple "single number" evaluations of student learning are crucial as they do not fit well outside formal educational setting. For instance, in the case of a MOOC student who enrolled in the course to gain basic familiarity with a particular subject domain, course completion or final course grade are misleading measures of learning success. Finally, the formal specification of analytics systems eases the replication of published research studies, as the specifics of the different learning settings are explicitly stated in the model operationalization and the definitions of student, task, and evidence models. As the field of learning analytics is maturing, the importance of study replications for developing sound and reliable empirical evidence is becoming increasingly important, and this model provides one potential approach for improving the replicability of published studies.

3 Assessing cognitive presence from online discussion transcripts

It is through others that we develop into ourselves.

— Lev Vygotsky, *The Genesis of Higher Mental Functions*

3.1 Introduction

THE focus of this chapter is on the use of transcripts of asynchronous online discussions for the assessment of student cognitive presence. Given the widespread use of asynchronous online discussions for supporting student interactions in online learning contexts (De Wever, Schellens, Valcke, & Van Keer, 2006; Luppicini, 2007), their transcripts provide valuable insights about students' learning activities. According to Henri (1992), discussion transcripts represent *"a gold mine of information concerning the psycho-social dynamics at work among students, the learning strategies adopted, and the acquisition of knowledge and skills."* (p. 118). In this chapter, we outline our research and contributions that utilized this "gold mine of information" for assessment of students' cognitive presence.

3.1.1 Brief overview of cognitive presence theoretical foundations

Before we explain and review the analytics system developed in this chapter, it is important to briefly examine the theory behind the cognitive presence construct as it forms the foundations of this thesis.

The origins of the cognitive presence can be traced back to the work of Dewey's (1910) who posited that the primary purpose of education is to develop reflective and higher-order critical thinking. In this process, the role of a critical community of inquiry is essential, as it enables students to challenge and question one another, demand reasons for beliefs, and diagnose misconceptions, ultimately resulting in a shared communal negotiation of meaning and co-construction of knowledge (Dewey, 1910; Lipman, 1991). For Dewey (1910), knowledge has the personal meaning which is co-constructed by the learners as they engage in discourse and their own reflective thinking. Learning is seen as inseparable from the social context in which it occurs and has two dimensions:

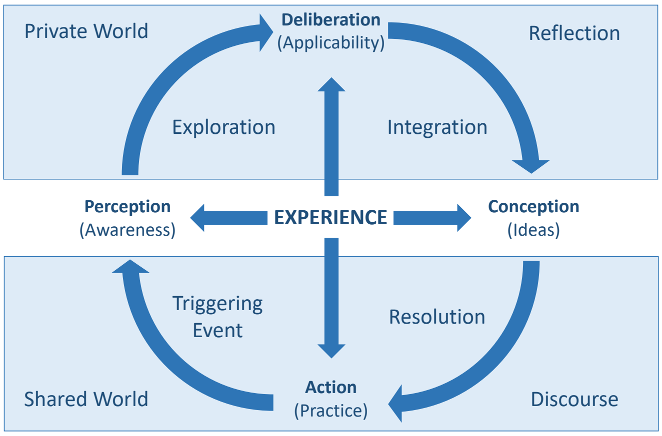


Figure 4. Practical Inquiry Model of cognitive presence (adopted from Garrison et al., 2001).

1) a psychological, individual dimension of reflective thought, and 2) a shared, social dimension of discourse and collaboration (Dewey, 1910).

Within the CoI model, cognitive presence construct is operationalized by the practical inquiry model (PIM) (Figure 4). The model defines cognitive presence as a cyclic process that consists of four phases (i.e., triggering event, exploration, integration, and resolution) of inquiry-learning during which students 1) identify problem, 2) explore ideas and information, 3) (co-)construct new knowledge, and 4) apply new knowledge to the originating problem. In this regard, the CoI model is a process model which views critical thinking as both a process and outcome of learning (Garrison, Anderson, & Archer, 2010). As shown in Figure 4, the practical inquiry model includes both reflective and social dimensions of learning, which is inspired by Dewey's (1910) views on the social nature of cognition and learning and views on practical inquiry (Lipman, 1991). The triggering event and resolution phases primarily involve discourse and social interactions around the real-world usability of existing and newly constructed knowledge, whereas exploration and integration phases primarily involve reflective thinking and personal construction of meaning.

To assess the levels of cognitive presence, Garrison et al. (2001) provide a coding instrument based on the quantitative content analysis (Krippendorff, 2003) of student discussion messages. As discussions represent the primary means of social interactions in the course (Garrison et al., 2001), they can be used to assess students' cognitive presence development as it is being expressed in the discourse. Looking at the whole messages as units of analysis, the CoI model defines a set of indicators that are indicative of particular theorized socio-cognitive processes associated with each phase of cognitive presence (Garrison et al., 1999). However, content analysis of student discussions is a time-consuming and labor-intensive activity (Donnelly & Gardner, 2011), typically requiring two or more trained and skillful coders. There are also a significant number of methodological challenges associated with its use (Rourke et al., 2001; Garrison et al., 2001; Strijbos, Martens, Prins, & Jochems, 2006; Riffe & Freitag, 1997), which hinder its wider usability by the practitioners. As a result, content analysis has been mostly used for research and retrospective purposes with a limited impact on the educational practice (Donnelly & Gardner, 2011).

3.1.2 Chapter overview

In this chapter, we present a novel learning analytics model for assessing cognitive presence based on student discussion transcripts. Using text mining (Aggarwal & Zhai, 2012) and natural language processing (Manning, Schütze, et al., 1999) techniques, we developed a text classification system which could be used to categorize discussion messages according to their levels of cognitive presence. The classification system makes use of metrics (features) extracted by the several existing tools for text analysis, such as LIWC (Linguistic Inquiry and Word Count) by Tausczik and Pennebaker (2010), and Coh-metrix by Graesser, McNamara, and Kulikowich (2011). The system also implements a set of novel metrics extracted through analysis of message content and discourse context (e.g., message's position in the discussion thread, semantic similarity to the previous/next message). Overall, the classification system obtained 70.3% accuracy and Cohen's $\kappa = 0.63$, indicating the potential of the adopted classification approach for the assessment of cognitive presence from the student discourse.

A significant contribution of the present work is that it provides means of assessing students' cognitive presence in a less labor-intensive and time-consuming way, which can potentially broaden the adoption of the CoI model by the practitioners and researchers. As CoI and other social-constructivist models require strong instructor presence to facilitate and guide student discussions, they are rarely used for courses with more than 30-40 students, primarily due to the limited capacity of instructors to attend to the high volume of interactions produced by the larger student cohorts (Anderson & Dron, 2010). In this regard, the classification system provides opportunities for instructors to attend to students discussions that most need instructional support and interventions.

Another major contribution of the present work is the more detailed data-informed operationalization of the different phases of cognitive presence with the set of extracted classification metrics. As the development of a classification system implicitly involves extraction of the associations between the levels of the dependent variable (i.e., phases of cognitive presence) and a set of classification metrics, it is possible to use this information to provide more detailed descriptions of each cognitive presence phase. For example, our results indicate that highest lexical diversity is expressed by the non-cognitive resolution messages, while resolution messages contain the largest number of content-related concepts. This and similar types of operations provide more detailed insights into the dynamics of cognitive presence and unique characteristics of each phase which in turn contribute to the current theoretical understanding of the cognitive presence construct.

3.2 Publication: Towards automated content analysis of discussion transcripts: A cognitive presence case

The following section includes the verbatim copy of the following publication:

> Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., and Siemens, G. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK'16)* (pp. 15–24). LAK '16. New York, NY, USA: ACM. doi:10.1145/2883851.2883950

Towards Automated Content Analysis of Discussion Transcripts: A Cognitive Presence Case

Vitomir Kovanović
School of Informatics
The University of Edinburgh
Edinburgh, UK
v.kovanovic@ed.ac.uk

Srećko Joksimović
Moray House School of Education
The University of Edinburgh
Edinburgh, UK
s.joksimovic@ed.ac.uk

Zak Waters
Queensland University of Technology
Brisbane, Australia
z.waters@qut.edu.au

Dragan Gašević
Moray House School of Education and School of Informatics
The University of Edinburgh
Edinburgh, UK
dgasevic@acm.org

Kirsty Kitto
Queensland University of Technology
Brisbane, Australia
kirsty.kitto@qut.edu.au

Marek Hatala
School of Interactive Arts and Technology
Simon Fraser University
Burnaby, Canada
mhatala@sfu.ca

George Siemens
LINK Research Lab
University of Texas at Arlington
Arlington, USA
gsiemens@uta.edu

ABSTRACT

In this paper, we present the results of an exploratory study that examined the problem of automating content analysis of student online discussion transcripts. We looked at the problem of coding discussion transcripts for the levels of cognitive presence, one of the three main constructs in the Community of Inquiry (CoI) model of distance education. Using Coh-Metrix and LIWC features, together with a set of custom features developed to capture discussion context, we developed a random forest classification system that achieved 70.3% classification accuracy and 0.63 Cohen's kappa, which is significantly higher than values reported in the previous studies. Besides improvement in classification accuracy, the developed system is also less sensitive to overfitting as it uses only 205 classification features, which is around 100 times less features than in similar systems based on bag-of-words features. We also provide an overview of the classification features most indicative of the different phases of cognitive presence that gives an additional insights into the nature of cognitive presence learning cycle. Overall, our results show great potential of the proposed approach, with an added benefit of providing further characterization of the cognitive presence coding scheme.

CCS Concepts

•Information systems → Clustering and classification; •Applied computing → E-learning; Distance learning;

LAK '16, April 25 - 29, 2016, Edinburgh, United Kingdom

ISBN 978-1-4503-4190-5/16/04...$15.00

DOI: http://dx.doi.org/10.1145/2883851.2883950

Keywords

Community of Inquiry (CoI) model, content analysis, content analytics, online discussions, text classification

1. INTRODUCTION

Online discussions are commonly used in modern higher education, both for blended and fully online learning [42]. In distance education, given the absence of face to face interactions, online discussions represent an important component of the whole educational experience. This is especially important for the social-constructivist pedagogies which emphasize the value of social construction of knowledge through interactions and discussions among a group of learners [2]. In this regard, the Community of Inquiry (CoI) model [23, 24] represents perhaps one of the best researched and validated models of online and distance education, focused on explaining important dimensions – also known as *presences* – that shape students' online learning experience.

The most commonly used approaches to the analysis of online discussion transcripts are based on the quantitative content analysis (QCA) [12, 54, 51, 16]. According to Krippendorff [37] content analysis is *"a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use"*[p18]. In the case of the study presented in this paper, contexts is online learning environments. QCA is a well defined research technique commonly used in social science research, and it makes use of specifically designed coding schemes to analyze text artifacts with respect to the defined research goals and objectives. For instance, the CoI model defines a set of coding schemes which are used by the educational researchers to assess the levels of three CoI presences.

In the domain of educational research, QCA of student discussion data have been mainly used for the retrospection and research after the courses are over without an impact on the courses' learning outcomes [53]. In the field of content analytics [36] – which focuses on building analytical models based on the learning content including student-produced content such as online discussion messages – there have been some attempts to automate some of those coding schemes. Most notable are the efforts of McKlin [44] and Corich et al. [11] on automation of the CoI coding schemes, which served as a starting point for our research in the area [35, 62]. One of the main challenges for automation of content analysis is the fact that the most important constructs from the educational perspective (e.g., student group learning progress, motivation, engagement, social climate) are latent constructs not explicitly present in the discussion transcripts. This means the assessment of these constructs requires human interpretation and judgment.

This paper presents the results of a study that explored the use of content analytics for automating content analysis of student online discussions based on the CoI coding schemes. We focused on automation of the content analysis of *cognitive presence*, one of the main constructs in the CoI model. By building upon the existing work in the fields of text mining and text classification and our previous work in this area [35, 62], we developed a random forests classifier which makes use of a novel set of classification features and provides a classification accuracy of 70.3% and Cohen's κ of 0.63 in our cross validation testing. In this paper, we describe the developed classifier and the adopted classification features. We also report on the findings of the empirical evaluation of the classifier and critically discuss the findings.

2. BACKGROUND WORK

2.1 The Community of Inquiry (CoI) model

The Community of Inquiry (CoI) model is a widely researched model that explains different dimensions of social learning in online learning communities [23, 24]. Central to the model are the three constructs, also known as *presences*, which together provide a comprehensive understanding of learning processes [23, 24]:

1) **Cognitive presence** which is the central construct in the CoI model and describes different phases of student knowledge construction within a learning community [24].
2) **Social presence** captures different social relationships within a learning community that have a significant impact on the success and quality of the learning process [50].
3) **Teaching presence** explains the role of instructors during the course delivery as well as their role in the course design and preparation [3].

The focus of this study is on the analysis of cognitive presence, which is defined by Garrison et al. [24] as *"an extent to which the participants in any particular configuration of a community of inquiry are able to construct meaning through sustained communication."*[p11]. Cognitive presence is grounded in the constructivist views of Dewey [14] and is "the element in this [CoI] model that is most basic to success in higher education" [23, p89]. Cognitive presence is operationalized by the practical inquiry model [24], which defines the following four phases:

1) **Triggering event**: In this phase, an issue, dilemma or problem is identified. In the case of a formal educational context, those are often explicitly defined by the instructors; however, they can also be initiated by the other discussion participants [24].
2) **Exploration**: This phase is characterized by the transition between the private world of reflective learning and the shared world of social construction of knowledge [24]. Questioning, brainstorming and information exchange are the main activities which characterize this phase [24].
3) **Integration**: In this phase, students move between reflection and discourse. The phase is characterized by the synthesis of the ideas generated in the exploration phase. The synthesis ultimately leads to the construction of meaning [24]. From a teaching perspective, this is the most difficult phase to detect from the discussion transcripts, as the integration of ideas is often not clearly identifiable.
4) **Resolution**: In this phase, students resolve the original problem or dilemma that started the learning cycle. In the formal educational setting, this is typically achieved through a vicarious hypothesis testing or consensus building within a learning community [24].

The CoI model defines its own multi-dimensional content analysis schemes [23, 24] and 34-item likert-scale survey instrument [5] which are used for the assessment of the three presences. The model has gained a considerable attention in the research community resulting in a fairly large number of replication studies and empirical validations (for an overview see [25]) including the studies about the interaction dynamics between the three presences [26]. In general, the model has been shown to be robust, and its coding scheme exhibits sufficient levels of inter-rater reliability for it to be considered a valid construct [25].

While the CoI model has been proven to be a very useful model for assessment of the social distance learning, there are several practical issues that still remain open. First, the use of the CoI coding schemes requires a substantial amount of manual work, which is very time consuming and requires trained coders. For example, to code the dataset used in this study, two experienced coders spent around 130 hours each to manually code 1,747 messages [22]. The coding process started with the calibration of the use of the coding scheme which was then followed by the independent coding, and finally reconciliation of the coding disagreements.

One major consequence of manual coding of messages in the CoI model is that it has been used mostly for research purposes and not for the real-time monitoring of students' learning progress and guiding instructional interventions. This is not unique to the CoI model and is very common with most of content analysis schemes used in education. The lack of automated content analysis approaches has been identified by Donnelly and Gardner [16] as one of the main reasons why transcript analysis techniques have had almost zero impact on educational practice. The development of the CoI survey instrument [5] is one attempt to eliminate, or at least to lessen the need for the manual content analysis of discussion transcripts. Still, the instrument is based on self-reported survey data, which makes it not so suitable for the real-time monitoring and guidance of student learning.

In order to enable for a broader adoption of the CoI model, the coding process needs to be automated and this is precisely the goal of the current study. While this study focuses on automation of coding online discussion transcripts for the levels of cognitive presence, a more general goal is to automate coding for all three presences, which would enable for a more comprehensive view of social learning phenomena and the development of more sophisticated social learning environments [60]. This in turn could be used by the instructors to inform their interventions leading to better achievement of learning objectives. From the standpoint of self-regulated learning research [7] – a major theory in contemporary education – in order to regulate their own learning effectively, learners need real-time feedback, which is an *"inherent catalyst"* for all self-regulated activities [7]. By providing learners with timely feedback on their own learning and the learning of their peers, they would be in a position to better regulate their own learning activities.

2.2 Automating Cognitive Presence Analysis

Several studies have investigated automating content analysis using the cognitive presence coding scheme. A study by McKlin [44] describes a system built using feed-forward, back-propagation artificial neural network that was trained on a single semester worth of discussion messages (N=1,997). The classification features were the counts of words in the one of the 182 different word categories as defined in the General Inquirer category model [52]. McKlin [44] also used a binary indicator whether a message is a reply to another message, as triggering events are more likely to be the discussion starters and thus not replies to other messages. Finally, McKlin [44] defined custom categories of words and phrases, which are thought to be indicative of the different phases of cognitive presence and included count of words in those categories as additional classification features. For example, "indicative words" category contains "compared to", "I agree", "that reminds me of", and "thanks" as it is hypothesized that integration messages would contain larger number of these phrases in order to connect the message with the previously given information. Unfortunately, these additional coding categories are very briefly described and thus is not possible to replicate them and evaluate their usability in future studies. McKlin's findings show that classification system overgeneralized the exploration phase and under-generalized the integration phase. Furthermore, given the very low frequency of messages in the resolution phase (i.e., $< 1\%$ and only 3 messages in total in their data set), the neural network developed by McKlin simply ignored the resolution category and never predicted the resolution phase for any message in the corpus. Overall, they reported Holsti's Coefficient of Reliability [30] of 0.69 and Cohen's κ of 0.31, which show some potential of the proposed approach with much room for improvement in order to reach reliability levels commonly found among two independent coders – usually Cohen's κ of at least 0.70 [28].

Following the work of McKlin [44], a study by Corich et al. [11] presented ACAT, a very general classification framework that can support any coding scheme besides cognitive presence which is also based on word count features. In order to use ACAT, users are required to provide a set of labeled training examples, which are used for training of classification models. Furthermore, as ACAT does not specify a particular set of word categories that are used as classification features, users are required to provide definitions (i.e., category name and list of words) that are used as classification features. Interestingly, the use of the ACAT system is also evaluated on the problem of coding cognitive presence of the CoI model. However, instead of classifying each message to one of the four phases of cognitive presence, Corich et al. [11] classified each *sentence* of each message to four cognitive presence levels. This poses some theoretical challenges as the CoI coding schemes are originally designed to be used for message-level content analysis. The dataset used by Corich et al. [11] consists of 484 sentences originating from 74 discussion messages and they report Holsti's coefficient of reliability of 0.71 in their best test case. However, given that their report did not provide sufficient details about the classification scheme used in terms of the specific indicators for each category of cognitive presence, nor did it discuss the types of features that were used for classification, it is hard to evaluate the significance of their results.

Besides the studies by McKlin [44] and Corich et al. [11], we should also mention our previous work in this domain. A study by Kovanović et al. [35] investigated the use of Support Vector Machines (SVMs) [59] classification for the automation of cognitive presence coding using a bag-of-words approach based on the N-gram and Part-of-Speech (POS) N-gram features. Using a 10-fold cross-validation, a classification accuracy of 0.41 Cohen's κ was achieved – which is higher than values reported in the previous studies [44, 11].

Several challenges related to the classification of online discussion massages based on cognitive presence were observed in our existing work [35]. First, the distribution of classes in the used dataset (i.e., phases of cognitive presence) was uneven, which is in agreement with the findings commonly reported in the literature [25]. This poses some challenges to the classification accuracy. This was already seen in the McKlin [44] study whose classifier completely ignored the resolution phase (as only three messages were coded as being in resolution phase). Secondly, the use of bag-of-words features (i.e., n-grams, POS n-grams, and back-off n-grams) creates a very large feature space (i.e., more than 20,000 features) relative to the number of classification instances (i.e., 1,747) which poses challenge of over-fitting. Next, the use of bag-of-words features makes the classification system highly domain dependent, as the space of bag-of-words features is defined based on the training set. For instance, a classification system trained on a introductory programming course would likely have a bigram feature `java programming` which is highly specific to a particular domain and would impede the performance of the classifier in other domains. Finally, given that each message belongs to a discussion and represents a part of the overall conversation, the context of the previous messages in the discussion thread is very important. For example, given the structure and cyclic nature of inquiry process, it is highly unlikely that a discussion would start with a resolution message, or that the first response to a triggering message will be an integration message [22]. These "dependencies" between discussion messages are not taken into the account when each message is classified independently of other messages in the discussion.

In order to address the challenge of isolated classification of discussion messages, Waters et al. [62] developed a structured classification system using conditional random fields (CRFs) [38]. This classifier does a prediction for the whole sequence of messages within a discussion, taking into the account orderings of messages within a discussion thread. Using a 10-fold cross-validation, the developed classifier achieved Cohen's κ of 0.48 which is significantly higher than 0.41 Cohen's κ reported by [35], showing a promise of the structured classification approach. However, there are still couple of unresolved issues which warrant further investigation. First of all, although the classification accuracy is improved, it is still far below the Cohen's κ of 0.7 which is considered a norm for assessing the quality of the coding in the CoI research community [28]. Secondly, CRFs are an example of black-box classification method [27] that are hard to interpret, which limits their potential use for *understanding* how cognitive presence is captured in the discourse.

3. METHOD

3.1 Data set

The dataset used in this study is the same dataset that was used in studies by Kovanović et al. [35] and Waters et al. [62]. The data comes from a masters level, and research-intensive course in software engineering offered through a fully online instructional condition at a Canadian open public university. The dataset consists of six offerings of the course between 2008 and 2011 with the total of 81 students that produced 1,747 discussion messages (Table 1). On average, each offering of the course had $\sim$ 13-14 students ($SD = 5.1$) that produced on average $\sim$ 291 messages, albeit with a large variation in the number of messages per course offer ($SD = 192.4$). The whole dataset was coded by the two expert coders for the four levels of cognitive presence enabling for a supervised learning approach. The inter-rater agreement was excellent ($percent\ agreement = 98.1\%$, Cohen's $\kappa = 0.974$) with a total of only 33 disagreements.

Table 2 shows the distribution of four phases of cognitive presence. In addition to the four categories of cognitive presence, we included the category "other", which is used for messages that did not exhibit signs of any phase of cognitive presence. The most frequent

Table 1: Course offerings statistics

	Student count	Message count
Winter 2008	15	212
Fall 2008	22	633
Summer 2009	10	243
Fall 2009	7	63
Winter 2010	14	359
Winter 2011	13	237
Average (SD)	13.5 (5.1)	291.2 (192.4)
Total	81	1,747

Table 2: Distribution of cognitive presence phases

ID	Phase	Messages	(%)
0	Other	140	8.0%
1	Triggering Event	308	17.6%
2	Exploration	684	39.2%
3	Integration	508	29.1%
4	Resolution	107	6.1%
	Average (SD)	349.4 (245.7)	20.0% (10.0%)
	Total	1,747	100%

messages were exploration messages (39% of messages), while the least frequent were the resolution messages (6% of messages). This large difference between the frequencies of the four phases was expected. It is consistent with the previous studies of cognitive presence [26], which found that a majority of students were not progressing to the later stages of integration and resolution. While there are various interpretations for this pattern, including the validity of the model, the design and expectations of the courses – i.e., not requiring students to move to those phases – seems to be the most compelling reason, as shown by its growing acceptance in the literature [25]. Psychologically, if students are going through the four phases of the practical inquiry model that underlies the cognitive presence construct, it does seem reasonable that students will spend more time exploring and hypothesizing different solutions, before they could come up with a final resolution [1, 22]. Moreover, as discussions were designed to occur between the third and the fifth week of the course, students did not typically move to the resolution phase this early in the course. Specifically, the discussions were organized to provide the students with opportunities to discuss ideas that would inform the individual research projects that they planned for the later stages of the course.

3.2 Feature Extraction

While the majority of the previous work related to text classification is based on lexical N-gram features (e.g., unigrams, bigrams, trigrams) and similar features (e.g., POS bigrams, dependency triplets), we eventually decided not to include N-gram and similar features described in the Kovanović et al. [35] study for several reasons. First of all, the use of those features inflates the feature space, generating thousands of features even for small datasets. This strongly increases the chances for over-fitting the training data. Secondly, the use of those features is also very “dataset dependent”, as data itself defines the classification space. Thus, it is hard to define a fixed set of classification features in advance, as the particular choice of words in the training documents will define what features are used for classification (i.e., what N-gram variables are extracted). Finally and most importantly, given that N-grams and other simple text mining features are not based on any existing theory of human cognition related to the CoI model, it is hard to understand what they might theoretically mean. Given that our goal is also to *understand* how cognitive presence is captured within discourse, we focused our work on extracting features which are strongly theory-driven and based on empirical studies. In total, we extracted 205 classification features which are described in the reminder of this subsection.

3.2.1 LIWC features

In this study, we used the LIWC (Linguistic Inquiry and Word Count) tool [57], to extract a large number of word counts which are indicative of different psychological processes (e.g., affective, cognitive, social, perceptual). Our previous research [33] showed that different linguistic features operationalized through the LIWC word categories offer distinct proxies of cognitive presence.

In contrast to extracting N-grams, which produce a very large number of independent features, LIWC provides us with exactly 93 different word counts which are all based on extensive empirical research [58, cf.]. LIWC features essentially “merge” related – and domain-independent – N-gram features together to produce more meaningful classification features. We used the 2015 version of the LIWC software package, which also provides four high-level aggregate measures of i) analytical thinking, ii) social status, confidence, and leadership, iii) authenticity, and iv) emotional tone.

3.2.2 Coh-Metrix features

For extraction of features for classification we also used Coh-Metrix [29, 45], a computational linguistics tool that provides 108 different metrics of text coherence (i.e., co-reference, referential, causal, spatial, temporal, and structural cohesion), linguistic complexity, text readability, and lexical category use. Coh-Metrix has been extensively used a large number of studies to measure subtle differences in different forms of text and discourse and is currently used by the Common Core initiative to analyze learning texts in K-12 education [45].

Coh-Metrix has been previously used in the domain of social learning to measure the student performance [17] and development of social ties [32, 34] based on the language used in the discourse. For example, a study by Dowell et al. [17] showed that characteristics of the discourse – as measured by Coh-Metrix – were able to account for 21% of the variability in the performance of active MOOC students. Students performed significantly better when then engaged in exploratory-style discourse, with the high levels of deep cohesion and the use of simple syntactic structures and abstract language. With the goal of the existing CoI content schemes to prescribe different indicators of important socio-cognitive processes in the discourse, the use of Coh-Metrix provides a valuable set of metrics that can be easily extracted and used for automation of the CoI coding schemes.

3.2.3 Discussion context features

Drawing on the study by Waters et al. [62], we also focused on incorporating more context information in our feature space. Thus, we included all features (except unigrams) which were used in the Waters et al. study. Those included:

- *Number of replies*: An integer variable indicating the number of replies a given message received.
- *Message Depth*: An integer variable showing a position of message within a discussion.
- *Cosine similarity to previous/next message*: The rationale behind these features is to capture how much a message builds on the previously presented information.
- *Start/end indicators*: Simple 0/1 indicator variables showing whether a message is first/last in the discussion.

As the CoI model – from the perspective of educational psychology – is a process model [25], students’ cognitive presence is viewed as being *developed over time* through discourse and reflection. Therefore, in order to reach higher levels of cognitive presence students need to either:i) construct knowledge in the shared-world through the exchange of a certain number of discussion messages, or ii) construct knowledge in the their own private world of reflective learning. Given the social-constructivist view of learning in the CoI model, we can expect that the distribution of messages exhibiting the characteristics of the different phases of cognitive presence will tend to change over time, as the students progress through those phases. Thus, we can expect that triggering and exploration messages will be more frequent in the early stages of the discussions, while integration and resolution messages will be more common in the later stages.

3.2.4 LSA similarity

Messages belonging to different phases of cognitive presence are characterized with various socio-cognitive processes [24]. The triggering phase introduces a certain topic in a tentative form, presenting a concept(s) that might not be completely developed, while the exploration phase further elaborates on various approaches to the inquiry initiated in the triggering phase. More precisely, the exploration phase introduces new ideas, divergent from the community, or even several contrasting topics within the same message [49]. On the other hand, the integration phase assumes a continuous process of reflection and integration, which leads to the construction of meaning from the introduced ideas [24]. Finally, the resolution phase presents explicit guidelines for applying knowledge constructed through the inquiry process [24, 49]. Based on these insights, we assumed that information presented in the various stages of the learning process might have an important influence on message comprehension. Still, given the differences among the learners and their learning habits, we did not expect this to be manifested as a general rule, but more as a slight tendency which would be useful in combination with the other classification features.

Following the approach suggested by Foltz et al. [20], we used LSA with the sentence as a unit of analysis to define a single variable `lsa.similarity`, which represents the average sentence similarity (i.e., coherence) within a message. As LSA determines the coherence based on the semantic relatedness between terms (i.e., terms that tend to occur in a similar context) [13], we first had to define a semantic space in which the similarity estimates are given. Having in mind that different discussions might relate to the different concepts, we decided to create a separate semantic space for each discussion. We identified the most important concepts from the first message in a discussion with a semantic annotation tool TAGME [19] and then each identified concept was linked to an appropriate Wikipedia page from which we extracted information about that concept [19]. Given that previous studies [55, 21] showed that Wikipedia can be used for estimation of semantic similarity between different concepts, we used information from the extracted pages to construct the semantic space on which LSA similarity of the concepts is calculated.

3.2.5 Number of named entities

Based on the work described in [47] and our previous study [35], we hypothesized that messages belonging to the different phases of cognitive presence would contain different count of named entities (e.g., named objects such as people, organizations, and geographical locations). The basis for this is taken from the definition of the cognitive presence construct [24]. Exploration messages are characterized by the brainstorming and exploration of new ideas, and thus, those messages are expected to contain more named entities than integration and resolution messages. Given the subject of the course in which the data for this study were collected, we extracted from each message a number of entities that are related to the computer science category of Wikipedia by using the DBPedia Spotlight annotation tool [46].

3.3 Data preprocessing

As the first step in our analysis, we addressed the problem of different number of messages in five classification categories (i.e., four phases of cognitive presence and "other"). The imbalance of different classes can have very negative effects on the results of the classification analyses [56]. Generally speaking, there are two possible ways of addressing this problem [8]: i) cost-sensitive classification, in which different penalties are assigned for misclassification of instances from different categories (higher penalties for smaller classes), and thus forcing the algorithm to put more emphasis on properly recognizing smaller classes; and ii) resampling methods, either by oversampling smaller classes, undersampling large classes, or through a combination of these two approaches. Given that cost-sensitive classification is used typically for two class problems ("positive" vs. "negative"), where correctly classifying one of the classes is the primary goal of the classifier (i.e., patients with a disease, fraudulent banking transaction), it makes sense to assign different misclassification costs as correctly identifying "negative" class is not important. However, in our case, we are equally interested in all five classes (four cognitive presence categories and the other messages), as they represent different phases in student learning cycles and it is not immediately clear whether misclassification of resolution messages is "worse" than misclassification of triggering event messages. Thus, in our study, we used resampling techniques and in particular a very popular SMOTE algorithm [9], which is a hybrid approach that combines oversampling the minority class with undersampling of the majority class.

One interesting property of SMOTE is that instead of simply resampling minority class instances – which would generate simple copies of the existing data points – it generates new *synthetic* instances which are "similar" to the existing instances but not exactly the same. For example, in n-dimensional feature space, for every data point ($X = \{f_1, f_2, ...f_n\}$) of the class C_i that is selected for resampling, SMOTE:

1) Find K (in our case five) nearest neighboring instances from the class C_i. As the distances between original C_i data points are known in advance, the list of K nearest neighbors for all instances in C_i class are calculated and stored in $N \times K$ matrix (where N is the number of data points in the C_i class).

2) Randomly picks one of the identified neighbors (Y).

3) Generates a new data point Z as:

$$Z = X + rand(0, 1) * Y$$

where $rand(0, 1)$ is a function returning a random number between 0 and 1.

Figure 1 shows the results of applying SMOTE to our dataset. As our original dataset consists of 1,747 messages, the class distribution would be uniform if each of the classes contained approximately 350 messages (i.e., $1,747/5 \sim 350$). Thus, we first user SMOTE oversampling procedure explained previously to generate additional 210, 42, and 243 instances of "Other", "Triggering", and "Resolution" classes, respectively. This increased the total number of messages in each of these three classes to 350 messages in total. We then undersampled messages in "Exploration" and "Integration" categories to create a smaller groups of also 350 messages. Hence, we removed 334 "Exploration" messages and 158 "Integration" messages, to produce smaller groups of also 350 messages in total. Overall, after applying SMOTE the new dataset consists of 1,750 messages, with each of the five categories of messages represented with exactly 350 messages.

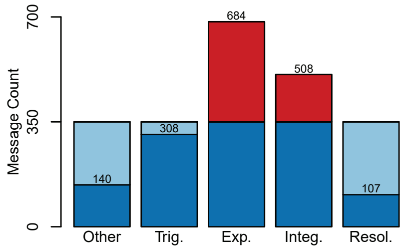


Figure 1: SMOTE preprocessing for class balancing. Dark blue – original instances which are preserved, light blue – synthetic instances, red – original instances which are removed.

Besides compensating for class imbalance problem, we also removed the two duplicate features that were provided by both LIWC and Coh-Metrix: i) the total number of words in a message, and ii) the average number of words in a sentence. We decided to remove LIWC values and use only the ones provided by Coh-Metrix. The primary reason for using Coh-Metrix features is consistency, as there are some small differences in how those two systems process corner cases (e.g., hyphenated words, interpunction signs) and given that Coh-Metrix provides additional set of metrics (e.g., number of sentences, number of paragraphs) we wanted to use consistent calculations for all of the included metrics.

3.4 Model Selection and Evaluation

To build our classifier, we used random forests [6], a state-of-the art tree-based classification technique. A large comparative analysis of 179 general-purpose (i.e., not domain-specific, offline, and unstructured) classification algorithms on 121 different datasets used in the previously published studies by Fernández-Delgado et al. [18] found that random forests were the top performing classification algorithm, only matched by Gaussian kernel SVMs. Random forests are ensemble tree-based method that combines bagging (bootstrap aggregating) with the idea of random-subspace to create a robust classification system which has low variance without increasing the bias [18]. Random forests work by creating a large number of trees and then the final prediction is decided using the majority voting scheme. Each tree is constructed on a different bootstrap sample (sub-sample of the same size with repetition) and evaluated on data-points that did not enter the bootstrap sample (in general, around one third of the training dataset size). In addition, each tree does not use the complete feature set, but has a *random* selection of N attributes (i.e., a subspace) which are then used for growing an individual tree without any pruning. Random forests are widely used technique that can handle large datasets with thousands of features.

It is important to note that random forests can also be used to measure importance of individual classification features. While importance of individual classification features might be calculated in many different ways [41], one popular measure is *Mean Decrease Gini (MDG)* which is based on the reduction in *Gini impurity* measure. Generally speaking, Gini impurity index measures how much the data points of a given tree node belong to the same class (i..e, how much they are "clean"). For every internal (split) node we can measure the decrease in Gini impurity, which shows how useful a given tree node is for separating the data (i..e, how much it reduces the impurity of the resulting groups of data). For random forests, MDG measure for a feature X_j is calculated as a mean decrease in Gini impurity of all tree nodes where a given feature X_j is used.

As there are two parameters used for configuration of random forests (i.e., `ntree` – number of trees constructed, and `mtry` – the number of randomly selected features), we used a cross-validation to select the optimal random forest parameters. As the performance

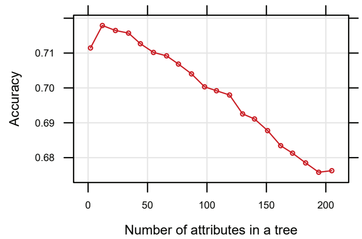


Figure 2: Random forest parameter tuning results.

of random forests typically stabilizes after a certain number of trees are built, we decided to build a large ensemble of 1,000 trees to make sure that convergence is reached. Thus, we focused on selecting optimal number of features used in every three (i.e., `mtry` parameter). We used a 10-fold cross validation and repeated it 10 times in order to reduce variability and get more accurate estimates of cross validated performance. In each run of the cross validation, we examined 20 different values for the `mtry` parameter: {2, 12, 23, 34, 44, 55, 66, 76, 87, 98, 108, 119, 130, 140, 151, 162, 172, 183, 194, 205}. The exact set of these values is obtained by using the `var_seq` function from R's `caret` package.

Before training and evaluating our classification models, we split data to 75% for model training and 25% for testing. We used stratified sampling, so that class distribution in both sub-samples is the same. We selected the best `mtry` value using the 10 repetitions of the 10-fold cross validation and then reported the classification accuracy of the best performing model on the testing data.

3.5 Implementation

We implemented our classifier in the R and Java programming languages using several software packages:

- for feature extraction we used Coh-Metrix [45, 29] and LIWC 2015 software packages [58],
- for developing random forest classifier, we used the `randomForest` R package [40],
- for running repeated cross validation and aggregating model performance, we used the `caret` R package [31],
- for running the SMOTE algorithm we used the Weka [63] Java package, and
- for calculation of LSA similarity measure, we used the Text Mining Library for LSA (TML)[1].

The complete dataset for the study and source code of the implementation is publicly available at `github.com/kovanovic/lak16_classification` repository.

3.6 Limitations

The major limitations of our approach are related to the size of our data set. Although we have six course offerings, they are all from the same course at a single university, and together with the particular details of adopted pedagogical and instructional approach they might potentially have an effect on the generalizability of our classification model. Thus, in our future work, we plan to test the generalization power of our classifier on a different dataset, which would preferably also account for other important confounding variables recognized in research of the CoI model such as subject domain [4], level of education (i.e., undergraduate vs. graduate) [26], and mode of instruction (blended vs. fully online vs. MOOC) [61].

4. RESULTS

[1]tml-java.sourceforge.net

Table 3: Random forest parameter tuning results

	mtry	Accuracy	Kappa
Min	194	0.68 (0.04)	0.59 (0.04)
Max	12	0.72 (0.04)	0.65 (0.05)
Difference		0.04	0.06

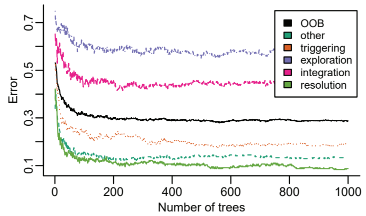


Figure 3: Best random forest configuration performance.

4.1 Model training and evaluation

Figure 2 shows the results of our model selection and evaluation procedure. The best classification accuracy of 0.72 ($SD = 0.04$) and 0.65 Cohen's κ ($SD = 0.05$) was obtained with `mtry` value of 12, which means that each decision tree takes into the account only 12 out of 205 features. The difference between the best- and worst-performing configurations was 0.06 Cohen's κ (Table 3), which suggest that parameter optimization plays an important role in the final classifier performance. Looking at the best performing configuration (Figure 3), we can see that the use of 1,000 trees in an ensemble resulted in reasonably stable error rates, with an average out-of-bag (OOB) error rate of 0.29, (i.e., an average misclassification rate for all data points in cases when they were non used in bootstrap samples). As expected, the highest error rates were associated with the undersampled classes (i.e., exploration and integration) and the smallest with the classes that were most heavily oversampled (i.e., resolution and "other").

Following the model building, we evaluated its performance on the hold-out 25% of the data. Our random forest classifier obtained 70.3% classification accuracy (95% CI[0.66, 0.75]) and 0.63 Cohen's κ which were significant improvements over 0.41 and 0.48 reported in Kovanović et al. [35] and Waters et al. [62] studies, respectively. Table 4 shows the confusion matrix obtained on the testing dataset. We can see that the most significant misclassifications are between exploration and integration messages which are hardest to distinguish. This is already witnessed in the [62] where most of the misclassifications were related to exploration and integration messages.

4.2 Variable importance analysis

Figure 4 shows the variable importance measures for all the 205 classification features. The median MDG score was 4.43, with the most of the features having smaller MDG scores, and only few features having very high MDG scores. Table5 shows the values of top 20 variables based on their MDG scores and their average values in

Table 4: Confusion matrix for the best performing model

	Predicted				
Actual	Other	Triggering	Explorat.	Integrat.	Resolut.
Other	**79**	2	2	2	2
Triggering	5	**67**	9	6	0
Exploration	9	15	**35**	27	1
Integration	2	2	23	**44**	16
Resolution	0	0	4	2	**81**

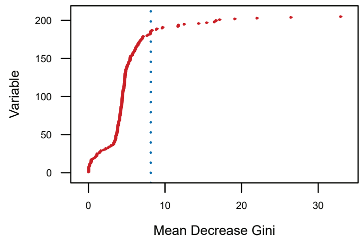


Figure 4: Variable importance by Mean Decrease Gini measure. Blue line separates top twenty features.

each class (i.e., cognitive presence phase). We can see that the most important variable was the `cm.DESCWC`, i.e., the number of words in a message; that is, the longer the message was, the higher the chance of the message was to be in the later phases of the cognitive presence cycle. Also, the number of paragraphs, number of sentences, and average sentence length showed similar trends, with higher values being associated with the later phase of cognitive presence.

The most important Coh-Metrix features were related to lexical diversity of the student vocabulary with the highest lexical diversity being displayed by "other" messages. Standard deviation of the number of syllables – which is an indicator of the use of words of different lengths – had the strongest association with the triggering event phase. In contrast, the givenness (i.e., how much of the information in text is previously given) had the highest association with the resolution phase messages. Finally, the low Flesch-Kincaid Grade level readability score and the low overlap between verbs used had the strongest association with "other" messages (i.e., messages without traces of cognitive presence).

The most important LIWC features were i) the number of question marks used, which was strongly associated with the triggering event phase, ii) the number of first person pronouns, which was highly associated with the other (i.e., non-cognitive presence) messages, and iii) use of money-related words, which is mostly associated with the integration and resolution phases.

Message context features also scored high, with message depth being higher for the later stages of cognitive presence, and highest for "other" messages. A similar trend was observed for similarity with the previous message, which was highest for the integration and resolution messages and lowest for the triggering event messages. In contrast, similarity with the next message and number of replies were highest for triggering events and lowest for the "other" messages. It is interesting to note that both LSA similarity and the number of named entities obtained high MDG scores. The number of named entities was the second most important feature and was highly associated with the later stages of the cognitive presence cycle. A similar trend was also observed for LSA similarity however, its importance was much lower.

5. DISCUSSION

Based on the testing results of the developed classifier, we can see that the use of the LIWC and Coh-Metrix features, together with a small number of thread-based context features could be used to provide reasonably high classification performance. The obtained Cohen's κ value of 0.63 falls in the range of "substantial" interrater agreement [39], and is just slightly below the 0.70 Cohen's κ which is the CoI research community commonly used as a threshold value for that is required before coding results are considered valid. We can also see that the parameter tuning plays an important role in optimizing the classifier performance, as the different classifier configurations obtained results different up to 0.05 Cohen's κ and 0.04% classification accuracy (Table 3).

Table 5: Twenty most important variables and their mean scores for messages in different phases of cognitive presence

				Cognitive presence phase				
#	Variable	Description	MDG*	Other	Triggering	Exploration	Integration	Resolution
1	`cm.DESWC`	Number of words	32.91	55.41 (61.06)	80.91 (41.56)	117.71 (67.23)	183.30 (102.94)	280.68 (189.62)
2	`ner.entity.cnt`	Number of named entities	26.41	13.44 (15.36)	21.67 (10.55)	28.84 (16.93)	44.75 (24.85)	64.18 (32.54)
3	`cm.LDTTRa`	Lexical diversity, all words	21.98	0.85 (0.12)	0.77 (0.09)	0.71 (0.10)	0.65 (0.09)	0.58 (0.09)
4	`message.depth`	Position within discussion	19.09	2.39 (1.13)	1.00 (0.90)	1.84 (0.97)	1.87 (0.94)	2.00 (0.68)
5	`cm.LDTTRc`	Lexical diversity, content words	17.12	0.95 (0.06)	0.90 (0.06)	0.86 (0.08)	0.82 (0.07)	0.78 (0.07)
6	`cm.LSAGN`	Avg. givenness of each sentence	16.63	0.10 (0.07)	0.14 (0.06)	0.18 (0.07)	0.21 (0.06)	0.24 (0.06)
7	`liwc.QMark`	Number of question marks	16.59	0.27 (0.85)	1.84 (1.63)	0.92 (1.26)	0.58 (0.82)	0.38 (0.55)
8	`message.sim.prev`	Similarity with previous message	16.41	0.20 (0.17)	0.06 (0.13)	0.22 (0.21)	0.30 (0.24)	0.39 (0.19)
9	`cm.LDVOCD`	Lexical diversity, VOCD	15.43	12.92 (33.93)	28.99 (50.61)	53.57 (54.68)	83.47 (43.00)	97.16 (28.95)
10	`liwc.money`	Number of money-related words	14.38	0.21 (0.69)	0.32 (0.74)	0.32 (0.75)	0.65 (1.12)	0.99 (1.04)
11	`cm.DESPL`	Avg. number of paragraphs sent.	12.47	4.26 (2.98)	6.37 (2.76)	7.49 (4.11)	10.17 (5.64)	14.05 (8.88)
12	`message.sim.next`	Similarity with next message	11.74	0.08 (0.14)	0.34 (0.40)	0.20 (0.22)	0.22 (0.24)	0.22 (0.23)
13	`message.reply.cnt`	Number of replies	11.67	0.42 (0.67)	1.44 (1.89)	0.82 (1.70)	1.10 (2.66)	0.84 (1.24)
14	`cm.DESSC`	Sentence count	11.67	4.28 (3.17)	6.36 (2.75)	7.49 (4.11)	10.17 (5.64)	14.29 (10.15)
15	`lsa.similarity`	Avg. LSA sim. between sentences	9.69	0.29 (0.27)	0.47 (0.23)	0.54 (0.23)	0.62 (0.20)	0.67 (0.17)
16	`cm.DESSL`	Avg. sentence length	9.60	11.88 (6.82)	13.62 (5.85)	16.69 (6.54)	19.36 (8.39)	21.73 (8.61)
17	`cm.DESWLsyd`	SD of word syllables count	8.92	0.98 (0.69)	1.33 (0.70)	0.98 (0.18)	0.97 (0.14)	0.97 (0.11)
18	`liwc.i`	Number of FPS* pronouns	8.84	4.33 (3.53)	2.82 (2.06)	2.37 (1.94)	2.51 (1.65)	2.19 (1.23)
19	`cm.RDFKGL`	Flesch-Kincaid Grade Level	8.29	7.68 (4.28)	10.30 (3.50)	10.19 (3.11)	11.13 (3.46)	11.99 (3.37)
20	`cm.SMCAUSwn`	WordNet overlap between verbs	8.14	0.38 (0.25)	0.48 (0.20)	0.51 (0.13)	0.50 (0.10)	0.47 (0.06)

MDG - Mean decrease Gini impurity index, **FPS** - first person singular

Given that the same dataset is used as in the [35] and [62] studies, it is possible to directly compare the results of the classification algorithms. The obtained Cohen's κ is 0.15 and 0.22 higher than the ones reported by Waters et al. [62] and Kovanović et al. [35], respectively. Furthermore, the resulting feature space is much smaller, with only 205 classification features in total, which is $\sim$100x smaller than the number of bag-of-words features extracted by Kovanović et al. [35] classifier. This limits the chances of over-fitting the training data and also improves the performance of the classifier. This is particularly important for the prospective use of the classifier in different subject domains, and pedagogical contexts.

Another important finding of this study is the list of important classification features. We see that a small subset of features is highly predictive of the different phases of cognitive presence, while a majority of the features have a much lower predictive power (Figure 4). It is interesting to notice that most of the discussion context features (except the discussion start/end indicators) obtained high importance scores, indicating the value in providing contextual information to the classification algorithm. In our future work, we will focus on investigation of the additional features that would provide even more contextualized information to the classifier.

It is important to notice that the list of the most important variables is aligned with the conceptions of cognitive presence in the existing CoI literature. If we look at the messages in the four phases of cognitive presence, we can see that the higher levels of cognitive presence are associated with messages that are i) generally longer, with more sentences and paragraphs, ii) adopt more complex language with generally longer sentences, iii) include more named entities (e.g., names of different constructs, theories, people, companies, and geographical locations) iv) have lower lexical diversity, v) occur later in the discussion, vi) have higher givenness of the information, higher coherence, and higher verb overlap, vii) use fewer question marks and first-person singular pronouns, viii) exhibit higher similarity with the previous messages, and ix) more frequently use money-related terms. Interestingly, the feature of the highest importance is also the simple word count implying that the longer the message, the more likely it is in the higher levels of cognitive presence cycle. This is also consistent with the findings of a previous study with the same dataset [33]. Joksimović et al. [33] found that word count was the only LIWC 2007 variable that yielded statistically significant differences among all four cognitive presence categories. This is not totally surprising as the similar findings are reported by essay grading studies who found that the strongest predictor of the final essay grade is the length of the essay [48].

Looking at the non-cognitive or "other" messages, we can see that they are characterized by the large lexical diversity. This is expected, as non-cognitive messages tend to be shorter (i.e., fewer words, paragraphs, and sentences) and more informal. Higher levels of lexical diversity are known to be associated with very short tests or texts of low cohesion [10]. As "other" messages often are not related to the course topic, they also tend to have a lower number of named entities, and lower givenness and verb overlap. Such messages also tend to adopt a simpler language, as indicated by the lowest scores on the Flesch-Kincaid grade level. "Other" messages also tend to occur more frequently near the end of the discussion, as indicated by their high values for `message.depth` feature and also more often are related to expression of personal information, as indicated by the highest values for the use of first-person singular pronouns. This is expected as many discussions would typically finish with students thanking each other for their contributions.

6. CONCLUSIONS

This paper has twofold contributions. First, we developed a classifier for coding student discussion transcripts for the levels of cognitive presence with a much higher performance (0.63 Cohen's κ) than previously reported ones [35, 62] in the studies with the same dataset. The performance of the developed classifier is in the range which is generally considered to be a substantial level of agreement [39]. We can see that the proposed approach, which is based on the use of Coh-Metrix, LIWC, and discussion context features, shows a great promise for providing a fully automated system for coding cognitive presence. The feature space that is used is also much smaller, which limits the chances for over-fitting the data and makes the developed classifier more generalizable to other contexts.

Secondly, we can see a particular subset of classification features that are very highly predictive of the different phases of cognitive presence. The most predictive feature is simple word count, which implies that the longer the message is, the higher the chances are for the message to display higher levels of cognitive presence. We also identified several additional features which are also highly predictive of the cognitive presence phase, in particular the number of named entities that are used (higher values are associated with integration and resolution phase) and lexical diversity (lower values are associated with "other" and triggering messages). We also see that features that provide information on the discussion context (i.e., similarity the with previous/next message, order in the discussion thread, and number of replies) are highly valuable and provide important information to the classification algorithm.

In our future work, we will focus on exploring additional features for improving the classification performance [43]. The study presented in this paper and our previous work [35] indicate that contextual features have a significant effect on classification accuracy and we will examine additional features of this kind. As our results reveal that the number of named entities has a significant effect on classification accuracy, and we will further explore similar features, such as concept maps [64], which would provide additional information about relationships between important concepts discussed in text-based messages. Finally, we will look at the different data preprocessing steps, including the use of the different algorithms for resolving the class imbalance problem. As we also observed that some of the students used direct quotes of other student messages which can cause problems for many of the text metrics that we used for classification, we will further examine the effects of the quotation on the final classification accuracy.

Finally, following the results presented in [15], we are exploring ideas for the development of a system that would – beside class labels – provide associated probabilities. Such a classifier could be used to develop a semi-automated classification system in which only one part of the data for which probabilities are sufficiently high would be automatically classified, and the rest would be manually classified. This would be advantageous as the *combined* desired accuracy of automatic-manual coding could be reached by setting a corresponding probability threshold. For achieving high levels of accuracy, a large majority of data would be classified automatically eliminating the large part of the manual work. Besides using it for coding discussion transcripts for research purposes, such system could be use, for example, to provide a real-time overview of the progress for a group of students and to point out the students for which an progress estimates are uncertain.

ACKNOWLEDGMENTS

This work was partially supported by the JISC [2] "Effective Learning Analytics" grant No.36 titled "Automated System for Cognitive Presence Coding".

REFERENCES

[1] Z. Akyol, J. B. Arbaugh, M. Cleveland-Innes, D. R. Garrison, P. Ice, J. C. Richardson, and K. Swan. A response to the review of the community of inquiry framework. *Journal of distance education*, 23(2), 2009. URL: http://www.ijede.ca/index.php/jde/article/view/630/884.

[2] T. Anderson and J. Dron. Three generations of distance education pedagogy. *The international review of research in open and distance learning*, 12(3):80–97, 2010. URL: http://www.irrodl.org/index.php/irrodl/article/view/890/.

[3] T. Anderson, L. Rourke, D. R. Garrison, and W. Archer. Assessing teaching presence in a computer conferencing context. *Journal of asynchronous learning networks*, 5:1–17, 2001. URL: http://auspace.athabascau.ca/handle/2149/725.

[4] J. B. Arbaugh, A. Bangert, and M. Cleveland-Innes. Subject matter effects and the community of inquiry (coi) framework: an exploratory study. *The internet and higher education*, 13(1):37–44, 2010.

[2] https://www.jisc.ac.uk/

[5] J. Arbaugh, M. Cleveland-Innes, S. R. Diaz, D. R. Garrison, P. Ice, J. C. Richardson, and K. P. Swan. Developing a community of inquiry instrument: testing a measure of the community of inquiry framework using a multi-institutional sample. *The internet and higher education*, 11(3–4):133–136, 2008.

[6] L. Breiman. Random Forests. *Machine learning*, 45(1):5–32, 2001.

[7] D. L. Butler and P. H. Winne. Feedback and self-regulated learning: a theoretical synthesis. *Review of educational research*, 65(3):245–281, 1995.

[8] N. V. Chawla, N. Japkowicz, and A. Kotcz. Editorial: special issue on learning from imbalanced data sets. *ACM SIGKDD explorations newsletter*, 6(1):1–6, 2004.

[9] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*:321–357, 2002. URL: https://www.jair.org/media/953/live-953-2037-jair.pdf.

[10] Coh-Metrix 3.0 indicies. http://cohmetrix.com/documentation_indices.html.

[11] S. Corich, K. Hunt, and L. Hunt. Computerised content analysis for measuring critical thinking within discussion forums. *Journal of e-learning and knowledge society*, 2(1), 2012. URL: http://www.je-lks.org/ojs/index.php/Je-LKS_EN/article/view/700.

[12] B. De Wever, T. Schellens, M. Valcke, and H. Van Keer. Content analysis schemes to analyze transcripts of online asynchronous discussion groups: a review. *Computers & education*, 46(1):6–28, 2006.

[13] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the american society for information science*, 41(6):391–407, 1990.

[14] J. Dewey. My pedagogical creed. *School journal*, 54(3):77–80, 1897.

[15] P. Dönmez, C. Rosé, K. Stegmann, A. Weinberger, and F. Fischer. Supporting CSCL with automatic corpus analysis technology. In *Proceedings of th 2005 conference on computer support for collaborative learning: learning 2005: the next 10 years!*, 2005, 125–134. URL: https://telearn.archives-ouvertes.fr/hal-00190638.

[16] R. Donnelly and J. Gardner. Content analysis of computer conferencing transcripts. *Interactive learning environments*, 19(4):303–315, 2011. URL: http://eprints.teachingandlearning.ie/3930/.

[17] N. Dowell, O. Skrypnyk, S. Joksimović, A. C. Graesser, S. Dawson, D. Gašević, P. d. Vries, T. Hennis, and V. Kovanović. Modeling Learners' Social Centrality and Performance through Language and Discourse. In *Proceedings of the 8th International Conference on Educational Data Mining (EDM 2015)*, 2015. URL: http://www.educationaldatamining.org/EDM2015/proceedings/full250-257.pdf.

[18] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim. Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research*, 15(1):3133–3181, 2014. URL: http://jmlr.org/papers/v15/delgado14a.html.

[19] P. Ferragina and U. Scaiella. Fast and accurate annotation of short texts with wikipedia pages. *Software, ieee*, 29(1):70–75, 2012.

[20] P. W. Foltz, W. Kintsch, and T. K. Landauer. The measurement of textual coherence with latent semantic analysis. *Discourse processes*, 25:285–307, 1998. URL: http://eric.ed.gov/?id=EJ589329.

[21] E. Gabrilovich and S. Markovitch. Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*. Morgan Kaufmann Publishers Inc., 2007, pp. 1606–1611. URL: http://dl.acm.org/citation.cfm?id=1625275.1625535.

[22] D. Gašević, O. Adesope, S. Joksimović, and V. Kovanović. Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *The internet and higher education*, 24:53–65, 2015.

[23] D. R. Garrison, T. Anderson, and W. Archer. Critical inquiry in a text-based environment: computer conferencing in higher education. *The internet and higher education*, 2(2–3):87–105, 1999.

[24] D. R. Garrison, T. Anderson, and W. Archer. Critical thinking, cognitive presence, and computer conferencing in distance education. *American journal of distance education*, 15(1):7–23, 2001.

[25] D. R. Garrison, T. Anderson, and W. Archer. The first decade of the community of inquiry framework: a retrospective. *The internet and higher education*, 13(1–2):5–9, 2010.

[26] R. Garrison, M. Cleveland-Innes, and T. S. Fung. Exploring causal relationships among teaching, cognitive and social presence: student perceptions of the community of inquiry framework. *The internet and higher education*, 13(1–2):31–36, 2010.

[27] L. Getoor. *Introduction to Statistical Relational Learning*. MIT Press, 2007. ISBN: 978-0-262-07288-5.

[28] P. Gorsky, A. Caspi, I. Blau, Y. Vine, and A. Billet. Toward a coi population parameter: the impact of unit (sentence vs. message) on the results of quantitative content analysis. *The international review of research in open and distributed learning*, 13(1):17–37, 2011. URL: http://www.irrodl.org/index.php/irrodl/article/view/1073.

[29] A. C. Graesser, D. S. McNamara, and J. M. Kulikowich. Coh-Metrix Providing Multilevel Analyses of Text Characteristics. *Educational researcher*, 40(5):223–234, 2011.

[30] O. R. Holsti. *Content analysis for the social sciences and humanities*. Addison-Wesley Reading, MA, 1969.

[31] M. K. C. f. Jed Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, t. R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, and C. Candan. *Caret: classification and regression training*. R package version 6.0-58, 2015. URL: http://CRAN.R-project.org/package=caret.

[32] S. Joksimović, N. Dowell, O. Skrypnyk, V. Kovanović, D. Gašević, S. Dawson, and A. C. Graesser. Exploring the Accumulation of Social Capital in cMOOC Through Language and Discourse. *Submitted*, 2015.

[33] S. Joksimović, D. Gašević, V. Kovanović, O. Adesope, and M. Hatala. Psychological characteristics in cognitive presence of communities of inquiry: A linguistic analysis of online discussions. *The internet and higher education*, 22:1–10, 2014.

[34] S. Joksimović, V. Kovanović, J. Jovanović, A. Zouaq, D. Gašević, and M. Hatala. What Do cMOOC Participants Talk About in Social Media?: A Topic Analysis of Discourse in a cMOOC. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, 2015, pp. 156–165.

[35] V. Kovanović, S. Joksimović, D. Gašević, and M. Hatala. Automated Content Analysis of Online Discussion Transcripts. In *Proceedings of the Workshops at the LAK 2014 Conference co-located with 4th International Conference on Learning Analytics and Knowledge (LAK 2014)*, 2014. URL: http://ceur-ws.org/Vol-1137/.

[36] V. Kovanović, S. Joksimović, D. Gašević, M. Hatala, and G. Siemens. Content Analytics: the definition, scope, and an overview of published research. In, *Handbook of Learning Analyitcs*, 2015.

[37] K. H. Krippendorff. *Content analysis: an introduction to its methodology*. Sage Publications, 2003.

[38] J. Lafferty, A. McCallum, and F. C. Pereira. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In *Proceedings of the eighteenth international conference on machine learning (ICML '01)*, 2001. URL: http://dl.acm.org/citation.cfm?id=655813.

[39] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.

[40] A. Liaw and M. Wiener. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002. URL: http://CRAN.R-project.org/doc/Rnews/.

[41] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts. Understanding variable importances in forests of randomized trees. In *Advances in neural information processing systems 26*, 2013, pp. 431–439. URL: http://media.nips.cc/nipsbooks/nipspapers/paper_files/nips26/281.pdf.

[42] R. Luppicini. Review of computer mediated communication research for education. *Instructional science*, 35(2):141–185, 2007.

[43] E. Mayfield and C. Penstein-Rosé. Using feature construction to avoid large feature spaces in text classification. In *Proceedings of the 12th annual conference on genetic and evolutionary computation*, 2010, 1299–1306.

[44] T. McKlin. Analyzing Cognitive Presence in Online Courses Using an Artificial Neural Network. PhD thesis. Georgia State University, College of Education, 2004.

[45] D. S. McNamara, A. C. Graesser, P. M. McCarthy, and Z. Cai. *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press, 2014.

[46] P. N. Mendes, M. Jakob, A. García-Silva, and C. Bizer. DBpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th international conference on semantic systems*, 2011, 1–8.

[47] J. Mu, K. Stegmann, E. Mayfield, C. Rosé, and F. Fischer. The ACODEA framework: developing segmentation and classification schemes for fully automatic analysis of online discussions. *International journal of computer-supported collaborative learning*, 7(2):285–305, 2012.

[48] E. B. Page and N. S. Petersen. The computer moves into essay grading: Updating the ancient test. *Phi delta kappan*, 76(7):561, 1995. URL: http://search.proquest.com/docview/218533317/abstract.

[49] C. L. Park. Replicating the Use of a Cognitive Presence Measurement Tool. *Journal of interactive online learning*, 8:140–155, 2, 2009. URL: http://www.ncolr.org/issues/jiol/v8/n2/replicating-the-use-of-a-cognitive-presence-measurement-tool#.VrVSebKUFhE.

[50] L. Rourke, T. Anderson, D. R. Garrison, and W. Archer. Assessing social presence in asynchronous text-based computer conferencing. *The journal of distance education / revue de l'éducation à distance*, 14(2):50–71, 2007. URL: http://eric.ed.gov/?id=EJ616753.

[51] L. Rourke, T. Anderson, D. R. Garrison, and W. Archer. Methodological issues in the content analysis of computer conference transcripts. *International journal of artificial intelligence in education (IJAIED)*, 12:8–22, 2001.

[52] P. J. Stone, D. C. Dunphy, and M. S. Smith. *The general inquirer: a computer approach to content analysis.* MIT press, 1966.

[53] J.-W. Strijbos. Assessment of (computer-supported) collaborative learning. *IEEE transactions on learning technologies*, 4(1):59–73, 2011.

[54] J.-W. Strijbos, R. L. Martens, F. J. Prins, and W. M. G. Jochems. Content analysis: what are they talking about? *Computers & education*, 46(1):29–48, 2006.

[55] M. Strube and S. P. Ponzetto. WikiRelate! Computing Semantic Relatedness Using Wikipedia. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*. AAAI Press, 2006, pp. 1419–1424. ISBN: 978-1-57735-281-5. URL: http://dl.acm.org/citation.cfm?id=1597348.1597414.

[56] P.-N. Tan, V. Kumar, and M. Steinbach. *Introduction to Data Mining*. Addison-Wesley Longman Publishing Co., Inc., 2005. ISBN: 0-321-32136-7.

[57] Y. R. Tausczik and J. W. Pennebaker. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of language and social psychology*, 29(1):24–54, 2010.

[58] Y. R. Tausczik and J. W. Pennebaker. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of language and social psychology*, 29(1):24–54, 2010.

[59] V. N. Vapnik. *Statistical learning theory*. Wiley-Interscience, 1998.

[60] J. Vassileva. Toward social learning environments. *IEEE transactions on learning technologies*, 1(4):199–214, 2008.

[61] N. Vaughan and D. R. Garrison. Creating cognitive presence in a blended faculty development community. *The internet and higher education*, 8(1):1–12, 2005.

[62] Z. Waters, V. Kovanović, K. Kitto, and D. Gašević. Structure matters: Adoption of structured classification approach in the context of cognitive presence classification. In *Proceedings of the 11th Asia Information Retrieval Societies Conference, AIRS 2015*, 2015.

[63] I. H. Witten, E. Frank, and M. A. Hall. *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann, 3rd ed., 2011.

[64] A. Zouaq and R. Nkambou. Building domain ontologies from text for educational purposes. *IEEE transactions on learning technologies*, 1(1):49–62, 2008.

3.3 Summary

In this chapter, we described a learning analytics system that provides an assessment of cognitive presence through the data analysis of course discussion transcripts. Similar to the current quantitative content analysis approach by (Garrison et al., 2001), the classification system developed in this chapter uses discussion transcripts to find indicators of the different phases of cognitive presence. However, unlike the manually time-consuming methods, the current system can be used to categorize the data for the duration of the courses and as such serve as a valuable support for instructors in monitoring student learning progress, and directing the attention to the discussions and students that need the most assistance.

With respect to the cognitive presence model described in Chapter two (Figure 4 of the Kovanović, Gašević, Hatala, and Siemens 2017 study included into this thesis), the described system represents one particular analytics implementation on the (top) assessment approach layer. The operationalization of the student and task models were given by the CoI model which were incorporated in the design of the course learning activities, while evidence model was operationalized using different messages extracted from online discussion transcripts and discourse context. As the cognitive presence assessment model takes into the account the design of learning activities and data sources in the specification of the analytics systems, the model can be used to adjust the analytics systems to different learning contexts and software platforms.

There are two major contributions of the work presented in this chapter. First, we developed a learning analytics system which can be used to classify student discussion messages based on message content and position within the threaded discussion. The analytics system shows a substantial improvement over the previously described systems (McKlin, 2004; McKlin, Harmon, Evans, & Jones, 2002; Corich et al., 2012; Kovanović, Joksimović, Gašević, & Hatala, 2014; Waters, Kovanović, Kitto, & Gašević, 2015) with an overall Cohen's $\kappa = 0.63$. Although still behind the level of agreement achieved by trained human coders, the achieved level of accuracy is sufficiently high enough to provide valuable insights to instructors regarding students' cognitive presence development. As such, the system could be used to provide actionable feedback to instructors upon which different interventions could be administered.

Another significant contribution of the work present in this chapter is the more detailed data-informed operationalization of cognitive presence and different phases of the practical inquiry model. In particular, through the analysis of the importance of the individual variables used in the classification process, we were able to identify the list of variables, most indicative of the different phases of cognitive presence. Figure 5 shows the graphical summary of the results presented in Table 5 of the Kovanović et al. (2016) study. We can see that for a majority of the variables, there is a clear increasing or decreasing trend across different phases of cognitive presence. The only exception are non-cognitive messages (coded as other) which in some cases (e.g., word count) have the lowest value, while in others (e.g., discussion position) have the highest value. For example, based on the results of this analysis, we can see that higher levels of cognitive presence are associated with:

- Longer messages, with more paragraphs and sentences,
- Higher language complexity,
- More coherent writing,
- Fewer first person singular pronouns,
- Longer average sentence length and their similarity,
- More content-related concepts mentioned,
- Lower lexical diversity (both on the content level and in general),
- Later discussion position,
- Fewer question marks, and
- More money-related terms.

As the cognitive presence coding scheme primarily focuses on high-level socio-cognitive processes indicative of the different phases of practical inquiry, the results presented in this chapter provide valuable insights into their association with the characteristics of students' written language. By looking at characteristics of students' messages as they progress through the phases of the practical inquiry cycle, the work presented in this chapter provides novel theoretical understanding of the dynamic nature of cognitive presence and its development.

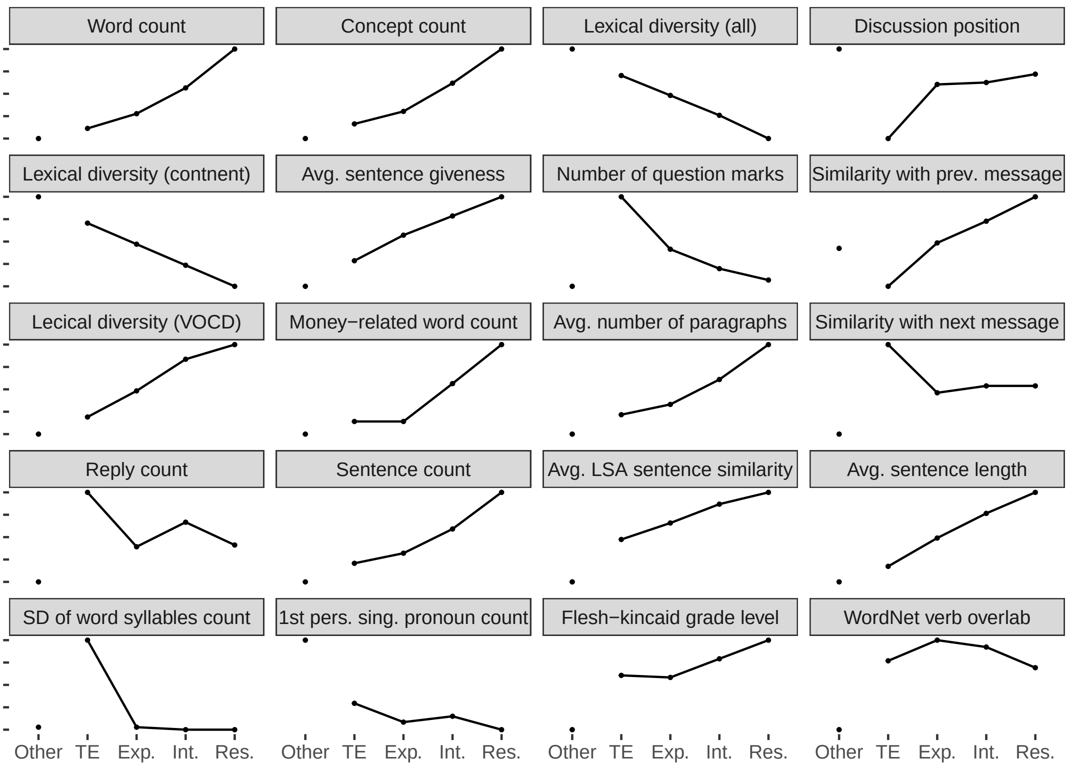


Figure 5. The most significant data-informed indicators of cognitive presence and their values across the messages in different cognitive presence phases.

4 Assessing cognitive presence using trace data

Absence of evidence is not evidence of absence.

— Carl Sagan, *The Demon Haunted World*

4.1 Introduction

The analysis of discussion transcripts provides valuable insights into students' cognitive presence development. The fact that students within a community of inquiry learn both through discourse and individual reflective learning activities means that there is a significant portion of cognitive and meta-cognitive activities which are not captured by the discussion transcripts. Indeed, this was recognized by Garrison et al. (2001) as *"observers view only that subset of cognitive presence that the participants choose to make visible in the conference. ... transcript of the conference is a significantly less-than-complete record of the learning that has taken place within the community of inquiry. Much work needs to be done, using triangulated measures supplemental to the conference transcript"* (p. 13). This chapter focuses on the use of digital trace data as "supplemental triangulated measures" to uncover the "undisclosed" part of cognitive presence that involves the personal and self-directed use of online learning environments. Given the recognized value of trace data for assessing students' self-regulating and reflective learning behaviors (Winne, 2006; Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007), in this chapter, we present a learning analytics system that makes use of the available trace data of students' interactions with the learning platform to identify key learning strategies adopted by the students in the course. As the identified strategies are indicative of students' metacognitive and self-regulatory behaviors, we also examine the differences in the cognitive presence of students who adopt different learning strategies.

4.1.1 Metacognition and self-regulation within communities of inquiry

It is widely recognized that students' success and learning experience, both face-to-face and online, depends to a large extent on a broad range of important individual factors such as their metacognitive abilities (Flavell, 1979), self-regulated learning (SRL) skills (Winne & Hadwin, 1998; Zimmerman, 2002; Azevedo & Aleven, 2013), self-efficacy (Bandura, 1977), and prior knowledge (Winne & Hadwin, 1998). As indicated by Akyol and Garrison (2011a), particularly related to critical thinking and inquiry-based learning is a student's ability to take control of their own learning processes, which is typically referred to as *metacognition*. Similarities between metacognition and critical thinking have been widely recognized (Sharma & Hannafin, 2004), with some (e.g., Martinez, 2006) even considering critical thinking to be a particular form of metacognition. Metacognitive control and monitoring are increasingly conceptualized as socially situated and constructed (Zimmerman, 2002; Larkin, 2009), with a substantial effect on students' participation in online discussions in a deep and meaningful way. Within contemporary educational psychology more broadly, metacognitive control and monitoring are viewed as integral components of a wider set of students' self-regulation of learning (Winne & Hadwin, 1998). As such, metacognitive control and monitoring are essential for self-directed information seeking and sense-making activities (Zimmerman, 2002) that characterize communities of inquiry (Shea et al., 2012).

Realizing the importance of learner's self-regulatory behaviors, several researchers (Shea & Bidjerano, 2010; Akyol & Garrison, 2011a; Shea & Bidjerano, 2012; Shea et al., 2012; Garrison & Akyol, 2013) focused on the analysis of students' metacognition, self-regulation, and self-efficacy and the ways in which it affects their online and blended learning experience and learning outcomes. Drawing on the work of Flavell (1979) and Pintrich and de Groot (1990), Akyol and Garrison (2011a) distinguish between static and dynamic facets of metacognition, and define metacognition within a community of inquiry as consisting of three dimensions: 1) *knowledge* of cognition (KC), 2) *monitoring* of cognition (MC), and 3) *regulation* of cognition (RC). Knowledge of cognition is viewed as a static element of metacognition which captures students' prior knowledge and motivation, while monitoring and regulation of cognition are seen as dynamic components of metacognition (Akyol & Garrison, 2011a). Similarly recognizing the mediating role of a student's self-regulation on educational experience and learning outcomes, Shea and Bidjerano (2010, 2012) propose *learning presence*, a new presence in the CoI model that captures self-regulation of learning (Shea & Bidjerano, 2010, 2012; Shea et al., 2012). Learning presence is conceptualized as consisting of the student's self-efficacy and effort regulation which are shown to be directly indicative of the student's learning success and reflection on the course content and learning tasks (Shea & Bidjerano, 2010, 2012). To assess students' self-regulated learning, both Akyol and Garrison (2011a) and Shea et al. (2012) provide early evaluations of two coding instruments for analysis of course discussion transcripts for the indicators of different self-regulatory behaviors.

From the direct importance for the work presented in this chapter is that metacognitive processes within the community of inquiry are seen as mediating elements between personal knowledge construction and social learning activities (Garrison & Akyol, 2013) and that as such, directly influence students' cognitive presence development (Shea & Bidjerano, 2010). In this regard, metacognitive monitoring *"includes the awareness and willingness to reflect upon the learning process. ...In practical terms, monitoring is facilitated by knowledge of practical inquiry."* (Akyol & Garrison, 2011a, p. 184) Similarly, metacognitive regulation is viewed as an interactive part of metacognition, and as a *"collaborative process where internal and external conditions are being constantly assessed."* (Akyol & Garrison, 2011a, p. 184). As such, to succeed in inquiry-based learning activities – which are to a large extent self-directed and discovery-oriented – students need to constantly engage in both self- and co-regulation of their learning through the use of different learning resources, tools, as well as through active participation in collaborative learning activities (Akyol & Garrison, 2011a; Garrison & Akyol, 2013).

Given the latent nature of self-regulatory processes, the most common approaches to assessing students' self-regulated learning are based on self-reported measures, such as the widely-used Motivated Strategies for Learning Questionnaire (MSLQ) by Pintrich, Smith, Garcia, and McKeachie (1991), Pintrich, Smith, Garcia, and Mckeachie (1993). However, with the advancements in educational technologies, there has been significant focus on the use of software systems as means of investigating complex study behaviors such as self-regulation and metacognition (Winne, 2006; Hadwin et al., 2007). The primary advantage of trace data is the ability to relatively easy and unobtrusively collect significant amounts of fine-grained data about learning tool use (known as trace data) in a systematic and objective manner (Azevedo & Aleven, 2013; Hadwin et al., 2007). This is especially important, given the evidence of substantial discrepancies between students' self-reports and trace data regarding important characteristics such as learning self-regulation (Hadwin et al., 2007) or goal orientation (Zhou & Winne, 2012).

4.1.2 Chapter overview

While the effects of student self-regulated learning and metacognition on student learning within traditional classroom context is widely recognized (Butler & Winne, 1995; Zimmerman, 2002; Flavell, 1979), there has been very limited research on the role of the student's self-regulation of learning within communities of inquiry (Akyol & Garrison, 2011a; Shea & Bidjerano, 2012). Although several approaches have been recently proposed, they are all based on either existing self-report instruments for the study of self-regulated learning (Shea & Bidjerano, 2010, 2012) or the quantitative content analysis of student discussion transcripts (Akyol & Garrison, 2011a; Shea et al., 2012; Garrison & Akyol, 2013). As such, they do not provide means for continuous and real-time assessment of student learning as it is happening which limits their practical usability by the educational practitioners to guide instructional interventions during student learning. Given that metacognitive monitoring and control have a direct effect on student learning strategy in terms of choice of different learning tools and resources, the analysis of trace data of students' use of available educational technology provides great potentials for assessing students learning self-regulation and metacognition.

This chapter focuses on examining the relationship between trace data about students' use of the available tools and resources and their levels of cognitive presence. Through a cluster analysis of students based on different measures of learning tool use, we identified a set of common learning strategies (i.e., clusters) of students' technology use and examined how the identified strategies differ with respect to cognitive presence. In the Kovanović, Gašević, Joksimović, et al. (2015) study included in this chapter, we referred to those learning strategies as "technology use profiles"; however, in this chapter we opted to use the term "learning strategy" to avoid possible unintentional confusions with learning styles. Our results indicate that there are significant differences in the level of cognitive presence among students who adopt different learning strategies. Moreover, our results indicate that students who adopted different learning strategies require specific interventions and instructional support in order to make changes towards strategies that are more effective. Learning strategies identified with learning analytics in combination with the classification system described in Chapter three can provide a comprehensive and in-depth assessment of student cognitive presence in a way which is not possible using just transcripts of student discussions.

4.2 Publication: Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions

The following section includes the verbatim copy of the following publication:

> Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., and Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, *27*, 74–89. doi:10.1016/j.iheduc.2015.06.002

Contents lists available at ScienceDirect

Internet and Higher Education

Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions

Vitomir Kovanović [a,*], Dragan Gašević [a], Srećko Joksimović [b], Marek Hatala [b], Olusola Adesope [c]

[a] University of Edinburgh, Edinburgh EH8 9YL, United Kingdom
[b] Simon Fraser University, Vancouver BC V5A 1S6, Canada
[c] Washington State University, Pullman, WA 99163, United States

ARTICLE INFO

Article history:
Accepted 10 June 2015
Available online 20 June 2015

Keywords:
Community of inquiry
Learner profiles
Clustering
Distance education
Human agency

ABSTRACT

This paper describes a study that looked at the effects of different technology-use profiles on educational experience within communities of inquiry, and how they are related to the students' levels of cognitive presence in asynchronous online discussions. Through clustering of students (N = 81) in a graduate distance education engineering course, we identified six different profiles: 1) task-focused users, 2) content-focused no-users, 3) no-users, 4) highly intensive users, 5) content-focused intensive users, and 6) socially-focused intensive users. Identified profiles significantly differ in terms of their use of learning platform and their levels of cognitive presence, with large effect sizes of 0.54 and 0.19 multivariate η^2, respectively. Given that several profiles are associated with higher levels of cognitive presence, our results suggest multiple ways for students to be successful within communities of inquiry. Our results also emphasize a need for a different instructional support and pedagogical interventions for different technology-use profiles.

1. Introduction

The importance of social interaction for reaching higher levels of learning is widely acknowledged in contemporary education (Anderson & Dron, 2010). Educational research offers many accounts of the benefits of social interaction on the development of skills such as critical thinking, creativity, and argumentation (Dawson, Tan, & McWilliam, 2011; Garrison, Cleveland-Innes, & Fung, 2010; Sawyer, 2006). Affordances of the modern (educational) technology enable for effective social interaction, information seeking, and knowledge building. More importantly, educational research offered approaches that can help design, facilitate, and direct an effective educational experience in communities and/or networks of learners. The Community of Inquiry (CoI) model (Garrison, 2011; Garrison, Anderson, & Archer, 1999; Garrison & Arbaugh, 2007) is a well-known framework in this context. By using qualitative and quantitative research methods, the research centered around the CoI model offered a remarkable amount of empirical evidence that explains an interplay of teaching, cognition, and socialization in communities of inquiry (Garrison, Cleveland-Innes, et al., 2010).

Although heavily dependent on educational technology, our review of the CoI literature revealed rather limited research that studied the relationships between learners' use of educational technology and the dimensions of the CoI model. The only study found in our literature review that focused on this issue was by Rubin, Fernandes, and Avgerinou (2013), and it investigated the association of learners' perceived value of educational technology affordances and perceived value of the core dimensions of the CoI model. However, the study of Rubin et al. used self-reports to gather students' perceived value of educational technology. In this paper, we propose that *learning analytics* (Buckingham Shum & Ferguson, 2012; Siemens & Gasevic, 2012) can: i) offer methods to advance understanding of the CoI model, especially in relation to learners' knowledge construction process and agency, ii) reveal how learners interact with educational technology in communities of inquiry, and iii) drive the development of new instructional approaches that can enhance educational experience for diverse subpopulations of learners that can emerge in communities of inquiry. More specifically, in this paper we report on the results of a study in which we:

1. Propose a *method* for identification of learner profiles – reflective of learners' agency about decisions making when selecting tools to study – based on trace data about their online learning activities performed in learning management systems.
2. Investigate the *effect* of the identified learner profiles on the development of cognitive presence – one of the three main dimensions of the CoI model – extracted from online discussion transcripts of a community of inquiry.
3. *Interpret* results in relation to instructional practice and existing theories on metacognition, motivation, and conceptions of and approaches to learning.

* Corresponding author.

http://dx.doi.org/10.1016/j.iheduc.2015.06.002

2. Theoretical background

2.1. The Community of Inquiry model

Built upon the social constructivist perspective to learning, the Community of Inquiry model is recognized by some as the most important model of e-learning today (Garrison & Arbaugh, 2007). The CoI model defines a community of inquiry as "*a group of individuals who collaboratively engage in purposeful critical discourse and reflection to construct personal meaning and confirm mutual understanding*" (Garrison, 2011, p. 2). The model describes a community of inquiry through the three interdependent dimensions, also known as *presences* (Garrison, 2007; Garrison, Anderson, & Archer, 2010; Kanuka, 2011):

1) *Cognitive presence* is a central dimension of the model that describes the learning phases from the initial practical inquiry to the eventual knowledge construction and problem solution (Garrison, Anderson, & Archer, 2001).
2) *Social presence* explains important social relationships among the members of the learning community and the social climate that contributes to the success of learning and attainment of the learning objectives (Rourke, Anderson, Garrison, & Archer, 1999).
3) *Teaching presence* is focused on the role of instructors in course design, organization, and delivery, and instructions that guide social and cognitive processes to desired learning outcomes (Anderson, Rourke, Garrison, & Archer, 2001).

This paper focuses on the study of cognitive presence which is defined as "*the extent to which the participants in any particular configuration of a community of inquiry are able to construct meaning through sustained communication*" (Garrison et al., 1999, p. 89). Cognitive presence proved to be a suitable instrument to assess critical thinking (Garrison et al., 2001), given that oral and textual communication (e.g., via discussion forums) have been shown to stimulate development of critical thinking skills. In essence, cognitive presence is a process model describing the development of higher-order thinking rather than individual learning outcomes (Akyol & Garrison, 2011b; Akyol et al., 2009). It is rooted in Dewey's (1910) social-constructivist views of learning and is operationalized through the practical inquiry model (Garrison et al., 2001) that defines four phases of inquiry learning cycle:

1) *Triggering event*: In this phase, a learning cycle is initiated by a problem or dilemma, which is in the formal educational setting typically introduced by the instructor.
2) *Exploration*: This phase is characterized by exploration, brainstorming, and other activities in which students gather information relevant to the problem or task at hand.
3) *Integration*: In this phase, after gathering an appropriate body of information, students synthesize and integrate different bits of information, while being selective and filtering out all irrelevant information.
4) *Resolution*: The last phase is the resolution of the original problem which is – in the context of formal education – typically achieved through vicarious actions and hypothesis testing. Very often resolution of the original problem initiates a new learning cycle with a new triggering event.

The research methods related to the community of inquiry include both: i) *qualitative methods* — primarily based on the use of quantitative content analysis of discussion transcripts and different coding schemes for the assessment of the three dimensions of the CoI model (Rourke & Anderson, 2004) and, ii) *quantitative methods* — primarily based on the CoI survey instrument, which was developed for measuring self-reported values of each of the three CoI dimensions (Garrison, Cleveland-Innes, et al., 2010). Both coding schemes (i.e., high inter-rater reliability) and the survey instrument (i.e., consistency and factor loadings) have been validated in a number of studies (Arbaugh et al., 2008; Gorsky, Caspi, Blau, Vine, & Billet, 2011; Rourke & Anderson, 2004).

Recent studies of the CoI model (Akyol & Garrison, 2011a; Garrison & Akyol, 2013; Shea & Bidjerano, 2010) highlight self-regulated learning (SRL) – a major theory of learning in contemporary educational psychology focusing on the role of metacognition in the learning processes (Bjork, Dunlosky, & Kornell, 2013) – as central for understanding the CoI model. As cognitive presence includes both self-reflection and collaborative knowledge co-construction (Garrison et al., 2001), "*metacognition mediates between reflection and action*" (Akyol & Garrison, 2011a, p. 186). In order to develop cognitive presence, students need to exercise critical thinking skills, which are primarily meta-cognitive in nature and require communicating one's thinking with others (Akyol & Garrison, 2011a). Garrison and Akyol's research showed that metacognition in a CoI could be characterized as "*complementary self- and co-regulation that integrates individual and shared regulation*" (Garrison & Akyol, 2013, p. 84). That is, participation in a community of inquiry affects their meta-cognitive monitoring and control. This is particularly done through the role of teaching presence whereby instructional design, facilitation, and direct instruction along with peer guidance are intrinsic components of metacognition in a community of inquiry.

2.2. Educational technology use and self-regulated learning

One of the central ideas in the modern educational psychology is that learners *do not acquire, but instead construct new knowledge* (Bjork et al., 2013; Winne, 2006; Winne & Hadwin, 1998). One of the major models which conceptually describes this process is self-regulated learning (SRL) (Bjork et al., 2013). It views knowledge construction as being developed through the use of different cognitive, physical, and digital tools to operate on raw materials to create the products of cognition. These products of cognition are evaluated with respect to standards that can be internal (e.g., efforts budgeted to online discussions) and external (e.g., grading policy for online discussions). Moreover, learners are viewed as human *agents* who constantly meta-cognitively: i) control their learning operations by evaluating their study tools, including decisions as to whether tools should be used and how to use the tools (Azevedo, 2005), and ii) monitor their learning progress by comparing the products of their learning with the predetermined learning goals.

As suggested by Winne (1982, 2006) and Perkins (1985), the knowledge construction and agency perspectives to learning have several important implications regarding the learners' use of tools. Typically, learning environments are designed to promote personalization and adaptiveness to the learners' needs (Azevedo, 2005). Still, studies indicate that many students do not make use of the available tools and resources in a way which will maximize benefits to the learning (Ellis, Marcus, & Taylor, 2005; Lust, Elen, & Clarebout, 2013a; Lust, Vandewaetere, Ceulemans, Elen, & Clarebout, 2011). Most of the available tools are underused by the majority of the students indicating the lack of awareness, knowledge, or motivation to use the available tools (Lust et al., 2013a). This is shown to be especially important in the complex, fully online environments given the self-directed nature of learning and the limited opportunities for the physical interactions among the students (Shen, Cho, Tsai, & Marra, 2013). Lust, Juarez Collazo, Elen, and Clarebout (2012) and Clarebout, Elen, Collazo, Lust, and Jiang (2013) indicate that for successful learning in the modern, complex learning environments learners:

1) need to be able to *recognize the opportunities* (e.g., tools or study tactics) that are available in the learning environment. Not all students have the needed meta-cognitive knowledge to recognize the provided learning opportunities (Clarebout et al., 2013) — e.g., the use of asynchronous online discussions for problem solving.
2) need to be able to *draw a connection* between the opportunity and their task at hand — e.g., that participation in the asynchronous

online discussions is beneficial for their learning.

3) need to be *meta-cognitively skillful enough* to be able to use the provided opportunity effectively. For example, to find relevant information in the course readings or the Internet, to articulate the information found in a meaningful way, and to finally integrate the information with the currently existing information in the course discussions.
4) need to be *motivated* to invest time and effort in using the opportunity and to meta-cognitively monitor and control the use of the opportunity in relation to their learning task. Likewise, students need to be comfortable with the types and extent of potential risks associated with the offered opportunity; For example, the potential misinterpretation of discussion contributions or the domination by a student or a group of students (Murphy & Coleman, 2004).

2.3. Technology-use profiles

In online and blended learning, one important aspect of student self-regulation of learning is the decision on if and if so, how to use the technology offered in a learning environment (Azevedo, 2005). Consistent with the research on self-regulated learning (Zhou & Winne, 2012), the studies of trace data recorded by learning management systems – typically based on a cluster analysis – showed that there were often several types of educational technology users (Lust et al., 2011, 2013a, 2013b; Wise, Speer, Marbouti, & Hsiao, 2013; Yen & Lee, 2011). It was also shown that patterns of educational technology use had different effects on learning outcomes of learners within the same course (Lust et al., 2013b).

In a blended learning environment, the study by Yen and Lee (2011) identified three groups of students based on their technology-use profiles: i) *technology-oriented* students who preferred mobile and web learning, and subsequently exhibited superficial problem solving abilities and general absence of planning and understanding, ii) *efficiency-oriented* students who were characterized by the efficient monitoring of their learning processes and generally better performance than the other two groups, and iii) *hybrid-oriented* students who did not have a preference for a particular instructional modality and mostly passively accepted information from the instructors. In a similar manner, the study by Lust et al. (2011) discovered three technology-use profiles in blended environments: i) *no-users* who did not make use of the available face-to-face tools (i.e., use of learning support and feedback sessions) and had a very limited use of the LMS, ii) *intensive* users that frequently used a majority of the tools available in the LMS, and iii) *incoherent* users who used only online tools and did not use the available face-to-face tools. With respect to the academic performance, both intensive and incoherent students had significantly higher academic performances in the course than those of the no-users.

One of the reasons for the observed differences was found to be students' *self-regulation* of the tool use (Lust et al., 2013a). Aligned with the findings in the field of self-regulated learning (Winne, 1982, 2006; Perkins, 1985), the study by Lust et al. showed that majority of the students regulated their learning; however, only 3% of them regulated in accordance with the course objectives. A majority of the students (59%) used a very limited set of available tools — indicating the lack of ability to regulate effectively their learning activities (Perkins, 1985).

Another construct that was explored by Lust et al. (2013b) is the students' achievement *goal-orientation* (Senko, Hulleman, & Harackiewicz, 2011), and it was found to be directly related to learning technology-use profiles of students. Generally, mastery goals focus on *gaining* competence while performance goals focus on *demonstrating* performance. In more recent studies, goal orientation is further distinguished along the emotional value that students give to their standards (approach vs. avoidance), resulting in four possible performance goal orientations: i) *mastery-approach*: focus on gaining skills and knowledge, ii) *mastery-avoidance*: focus on avoiding skill decline and learning failures, iii) *performance-approach*: focus on performing better than peers, and iv) *performance-avoidance*: focus on avoiding performing worse than peers (Senko et al., 2011). What is of the direct importance for the current study are the findings of connection between i) mastery goal orientation and active tool use, and ii) performance goal orientation and selective tool use (Lust et al., 2013b). Furthermore, the students did not differ in terms of their perceived value of the provided tools, and in some cases – such as practice quizzes – no-users even had a significantly more positive opinion on the tool usefulness despite the fact that they never used them.

The notion of *approaches to learning* (i.e., deep vs. surface) (Trigwell & Prosser, 1991) is another important construct in educational research which was shown to have a significant impact on the learning outcomes, particularly related to student technology use. Research findings indicate connection between deep learning approaches and higher success of learning (Trigwell & Prosser, 1991) and mastery goal orientation (Phan, 2008). With this in mind, student participation in online discussions was analyzed by Wise et al. (2013), and their study revealed that performance-avoidance was directly related to low cognitive engagement. They identified four distinctive profiles of participation different primarily in terms of the *breadth*, *depth*, and *temporal continuity* of discussion participation and the amount of student *reflection* (Wise et al., 2013). Thus, Wise et al. recommend that instructors need to take into the account differences among their students, especially in relation to their goals and approach to learning. Their findings are consistent with the results of Bliuc, Ellis, Goodyear, and Piggott (2010), who showed the association between surface approaches to learning and fragmented *conception of learning* in discussions (i.e., discussions were considered a way of getting a correct answers fast rather than a way of deepening their broader knowledge), and also cohesive notion of learning in discussions and deep approaches to learning.

From a more holistic perspective, Valle and Duffy (2009) consider the very own idea of *student freedom* in distance education as a challenge for many students. As a result, a key to success is seen in the effective managing of the learning demands associated with the freedom of distance education (Valle & Duffy, 2009). Valle and Duffy in their study found similar three types of technology-use profiles: i) mastery-oriented, who were shown to possess the greatest amount of relevant background experience and were willing to put substantial effort into learning, ii) task-oriented, who had the overall lower levels of effort and were spending the minimal required time, and iii) minimalist, who also put less effort like the task-oriented students, and were also found to prefer working in groups rather than self-paced — indicating a need for more motivation-related support. Their results suggest that the ways in which learners work might be a good indicator of their commitment to learning. Still, in terms of their success, both Wise et al. (2013) and Valle and Duffy (2009) did not find any significant difference among the students which suggest that many different approaches might be successful in terms of academic performance.

2.4. Educational technology use within the communities of inquiry

Educational technology is the major enabler of communities of inquiry. While the effects of technology use on learning and factors of technology acceptance received considerable research attention (McGill & Klobas, 2009), research of the effects of educational technology affordances on the three dimensions of the CoI model is hardly reported in the literature published to date. To our knowledge, a recent paper by Rubin et al. (2013) reports the first study that tried to shed some light on this important issue. By using the three dimensions of the CoI model and self-reports of the use of a selected set of LMSs, they found several patterns of how educational technology relates to the CoI model. Of direct importance for the study reported in this paper is the finding that the level of self-reported cognitive presence is predicted by the self-reported ease of communication provided by the LMS and the self-reported amount of online reading materials, while the self-reported ease of finding information was marginally significant. These findings were expected, as for effective participation in a community of inquiry (e.g., to achieve a high level of cognitive presence), not only did learners participate in online discussions, but rather their learning involved a number of activities such as seeking and reading learning materials, modeling their knowledge through quizzes, and completing course assignments.

Existing research on technology use and self-regulated learning offers a number of theoretical, methodological, and empirical accounts warranting future research on the relationships between educational technology use and the CoI model. It is widely accepted that external and internal conditions play an important role in regulating students' approach to study (Winne & Hadwin, 1998). Studies by Garrison and Cleveland-Innes (2005) and Akyol and Garrison (2011b) looked at learning approaches within communities of inquiry and showed that specific forms of teaching presence – such as instructional leadership in facilitation, direct instruction, and appropriate course structure – already have a positive effect on the promotion of deep approaches to learning, and thus, establishing and sustaining of high levels of cognitive presence.

With respect to the adopted methodological approaches, one of the primary means of studying technology use, student agency, and self-regulation of learning is through the self-reported data (Winne & Jamieson-Noel, 2002). The previously mentioned studies by Lust et al. (2013b), Bliuc et al. (2010), and Valle and Duffy (2009) are some of the examples. However, self-reports are not the most reliable instrumentation to study the effects of education technology on learning processes and outcomes, given the biases and inaccuracy associated with understanding metacognition. For example, Winne and Jamieson-Noel showed that learners tended to overestimate their use of study tools in a learning software. Although self-reports offer valuable insights into learners' perception of learning, previous research found that learners often self-report "*biased information arising from incomplete and reconstructed memories plus subjective and implicit theories of the mental processes involved*" (Zhou & Winne, 2012, p. 414). In order to overcome some of the challenges of self-reported measures, the use of more objective measures – such as students' trace data (Zhou & Winne, 2012, p. 414) – is often recommended (Gonyea, 2005).

3. Research questions

From the studies presented in Section 2.3 we can see a strong evidence supporting the difference among students in terms of their technology use, and the importance of students' goal-orientation, self-regulation, and approaches to learning on shaping technology-use profiles. From the existing research studies (Bliuc et al., 2010; Lust et al., 2011, 2013a, 2013b; Valle & Duffy, 2009; Wise et al., 2013; Yen & Lee, 2011), several reoccurring technology-use profiles can be seen: i) a group of students with lower activity levels (i.e., no-users and minimalist users) typically associated with lower levels of meta-cognitive capabilities for self-regulation, surface approaches to learning, and performance–goal orientation, ii) a group with very high levels of activity (i.e., intensive and mastery oriented-users) who had deep approaches to learning and mastery-goal orientation, and iii) a group of selective users (i.e., incoherent, selective, limited, and efficiency-oriented users) that typically exhibited performance goal-orientation and higher of learning self-regulation — although for most of the time not in a desirable way.

Our hypothesis is that we will find the same or very similar technology-use profiles within the communities of inquiry. Still, given the social-constructivist view of learning in communities of inquiry, we are interested in how this particular context affects the hypothesized profiles. Thus, our first research question is:

Research question 1. *What are the main technology-use profiles within communities of inquiry? How does the collaborative nature of learning within communities of inquiry affects the theorized technology use profiles? Are there any CoI-specific learning technology-use profiles not previously identified, and if so, how they can be explained in terms of the students' self-regulation of learning, goal-orientation and approaches to learning?*

In this paper, we build on the existing research of the effects of technology-use (Rubin et al., 2013) and self-regulation (Akyol & Garrison, 2011a; Garrison & Akyol, 2013) on the learning success within communities of inquiry. More precisely, we explore effects of individual, internal regulation – as evident through the different profiles of technology use – on the development of the cognitive presence. Building on the suggested relationship between approaches to learning and technology-use profiles, we investigate the relationship between the CoI model and student approaches to learning, as indicated by the observed technology-use profiles. Thus, our second research question is:

Research question 2. *How are the profiles of learning technology use related to the development of cognitive presence in communities of inquiry and which profiles have the strongest effect on cognitive presence?*

Given the existing evidence of the effects of approaches to learning and goal orientation on the development of deep critical thinking skills (Bliuc et al., 2010; Entwistle, 2009; Lust et al., 2013b; Phan, 2008; Trigwell & Prosser, 1991; Wise et al., 2013), we expect to find differences in terms of the students' development of cognitive presence and ultimately on the success of learning. In this study, we focus on the cognitive presence; however in the future studies we will also examine the effects on the academic performance — as operationalized through the final course grades. From the practical perspective, this research question seeks to provide: i) insights which can be potentially used for adaptation of the provided feedback to learners based on their technology-use profiles, and, ii) a guide for instructors that can help them to define specialized and eventually more effective instructional interventions targeting students with specific styles of educational technology use.

4. Methods and materials

4.1. Course

4.1.1. Course organization

The data for this study originated from a thirteen week long, master level course offered through a fully online instructional condition at a Canadian public university. The course is research intensive and focuses on understanding of the current research trends and challenges in the area of software engineering field. To successfully finish the course, students were expected to complete several activities including four tutor marked assignments (TMAs):

- *TMA1* (15% of the final grade, submitted during weeks 3–5): The students are expected to: i) select and read a peer-reviewed paper on a course topic, ii) prepare a short video presentation that summarizes information presented in the paper and provides a critical review of the paper and iii) initiate a new discussion about the paper with other students. This assignment is primarily factual, and focuses on presenting particular challenges in a software engineering field.
- *TMA2* (25% of the final grade, submitted at the end of week 6): The students were required to write a literature review paper (5–6 pages in the ACM proceedings format) on a selected topic in software engineering. The marking scheme for this assignment was as follows: i) 80% of the grade was given based on two double blind peer reviews (35% of the grade each) and instructor review (30% of the paper grade), and ii) 20% was given by the instructor based on the quality of provided peer-review comments. This assignment has strong focus on building conceptual understanding of a particular research problem in software engineering field.
- *TMA3* (15% of the final grade, submitted at the end of week 9): Students were required to answer six questions (400–500 words per question) related to course readings that were designed to

demonstrate critical thinking and synthesis skills. The focus of this assignment is also on conceptual knowledge and analysis and evaluation of the existing solutions of a given research problem.
- *TMA4* (30% of the final grade, submitted at end of the course): In the final assignment, students worked in small groups (2–3 students) on a selected software engineering topic. The main outcome was a project report and all developed software artifacts (e.g., models and source code) that were then marked by the instructor. This assignment has a particularly procedural focus on building practical skills related to the selected research topic and evaluation of proposed solutions.
- *Course participation* (15% of the final grade): The course had a particular focus on stimulating productive online discussions and the students were expected to actively participate in course discussions.

4.1.2. Dataset

The data consisted of the 6 offerings of the described course (Winter 2008, Fall 2008, Summer 2009, Fall 2009, Winter 2010, Winter 2011) with a total of 81 students with an average cohort size of 13.5 students (SD = 5.1). The slightly larger variations in cohort sizes were due to the course under study not being a mandatory course, but a part of the group of 11 'core' courses, and university regulations required students to complete three core courses in their master's degree programs.

The course was offered through the Moodle LMS,[1] which hosted all the readings, assignments and student discussion boards. The trace data was obtained by an automated extraction process from the Moodle's PostgreSQL database and consisted of almost 200,000 log records of different student activities. In these six offerings, the students posted 1747 messages in total which – together with the LMS trace data – represented the main data source for this study. The numbers of students and messages in each course is shown in Table 1.

In order to measure the levels of cognitive presence, all 1747 messages from online discussion forums were coded using the CoI coding instrument described in Garrison et al. (2001). All messages were coded by two human coders and they achieved an excellent coding agreement (Cohen's Kappa = 0.97), disagreeing in less than 2% of the messages (i.e., total of 32 messages). In those cases, the disagreements were resolved through the discussion between the coders. The results of the coding are shown in Table 2.

4.2. Measurement instrument

In order to identify technology-use profiles, thirteen variables based on students' use of LMS were extracted (Table 3), similarly to the work of Lust et al. (2011, 2013a, 2013b) and Valle and Duffy (2009). We extracted count and time-on-task variables, focusing only on LMS activities that students were expected to use given the particular course design. For most of the activities, both counts and time-on-task were extracted, while for some activities only count measures were extracted, as the notion of time-on-task was not meaningful (e.g., searching discussion boards). Table 3 shows that the extracted variables can be divided into two groups: variables related to the static course content (reading resources and assignments) and variables related to online discussions.

With respect to the outcome variables, we used the counts of messages in the phases of cognitive presence that were collected through a quantitative content analysis using CoI's cognitive presence coding scheme (Garrison et al., 2001) which is described in detail in Section 4.1.1. Hence, for each student five outcome measures were extracted, four corresponding to the four phases of cognitive presence and one corresponding to the messages without traces of cognitive presence (coded as other). Typically, non-cognitive (i.e., other) messages included messages serving purely social purposes, such as acknowledging someone else's message.

[1] http://moodle.org

Table 1
Course offering statistics.

	Student count	Message count
Winter 2008	15	212
Fall 2008	22	633
Summer 2009	10	243
Fall 2009	7	63
Winter 2010	14	359
Winter 2011	13	237
Average (SD)	13.5 (5.1)	291.2 (192.4)
Total	81	1747

4.3. Pre-processing clickstream data

As recorded trace data is mainly a stream of actions together with occurrence timestamps, the first step in our analysis was to preprocess trace data to extract count and time-on-task variables. Count measures were extracted by simply counting for each action the number of times that it was performed by each student, while time-on-task variables were calculated from the time differences between the logged actions. This is the typical approach that has been extensively used in other similar studies (Lust et al., 2011, 2013a, 2013b; Valle & Duffy, 2009; Wise et al., 2013), as well in many Learning Analytics and Educational Data Mining studies (Macfadyen & Dawson, 2010; Morris, Finnegan, & Wu, 2005; Romero, Ventura, & García, 2008). The primary assumption – which is commonly done in time-on-task estimation (Valle & Duffy, 2009) – is that time between two logged events is spent on a particular learning activity.

One particular challenge of this approach that has been already identified by Wise et al. (2013) and Valle and Duffy (2009) is the detection of time when a user has left the system. Even though LMSs have a logout button, a great majority of students do not use it and simply close their web browser window. Therefore, to prevent from severely overestimating time-on-task measures, durations of last activities for each study session (i.e., activities that were followed by a login action) were estimated as the student's average time for that particular activity. Finally, as sometimes students would just leave the browser window open for an extended period of times (and thus their next study session does not start with the login action), an upper limit of the duration of each activity was set to 1 h, similarly to the work of Valle and Duffy.

4.4. Clustering

For the discovery of students' technology-use profiles cluster analysis techniques and the popular agglomerative hierarchical clustering algorithm were adopted (Hastie, Tibshirani, & Friedman, 2013). Much like Wise et al. (2013) and Valle and Duffy (2009), we used Ward's merging procedure and Euclidean distance measure (Hastie et al., 2013). As some of the variables are counts and some are time durations, similarly to the work of Valle and Duffy all variables were first standardized in order to enable their equal weighting. Finally, each cluster was

Table 2
Cognitive presence coding.

ID	Phase	Messages	(%)
0	Other	140	8.01%
1	Triggering event	308	17.63%
2	Exploration	684	39.17%
3	Integration	508	29.08%
4	Resolution	107	6.12%
	All phases	1747	100%

Table 3
Extracted features.

#	Type	Code	Name	Description
1	Clustering variables (content)	ULC	UserLoginCount	Total number of times student logged into the system.
2		CVC	CourseViewCount	Total number of times student viewed general course information.
3		AVT	AssignmentViewTime	Total time spent on all course assignments.
4		AVC	AssignmentViewCount	Total number of times student opened one of the course assignments.
5		RVT	ResourceViewTime	Total time spent on reading the course resources.
6		RVC	ResourceViewCount	Total number of times student opened one of the course resource materials.
7	Clustering variables (discussions)	FSC	ForumSearchCount	Total number of times student used search function on the discussion boards.
8		DVT	DiscussionViewTime	Total time spent on viewing course's online discussions.
9		DVC	DiscussionViewCount	Total number of time student opened one of the course's online discussions.
10		APT	AddPostTime	Total time spent on posting discussion board messages.
11		APC	AddPostCount	Total number of the discussion board messages posted by the student.
12		UPT	UpdatePostTime	Total time spent on updating one of his discussion board messages.
13		UPC	UpdatePostCount	Total number of times student updated one of his discussion board messages.
1	Outcome variables	TEC	TriggeringEventCount	Number of posted *triggering event* messages.
2		EC	ExplorationCount	Number of posted *exploration* messages.
3		IC	IntegrationCount	Number of posted *integration* messages.
4		RC	ResolutionCount	Number of posted *resolution* messages.
5		OC	OtherCount	Number of posted non-cognitive (*other*) messages.

summarized by calculating the cluster's *centroid*, which represented the mean values of all cluster members across all clustering variables.

4.5. Statistical analysis

For assessing the difference between student clusters a multivariate analysis of variance (MANOVA) (Tabachnick & Fidell, 2007) was used. To validate the difference between the discovered clusters a MANOVA model with cluster assignment as a single independent variable and thirteen clustering variables (Table 3) as the dependent measures was constructed, similarly to the work of Lust et al. (2011, 2013a, 2013b). To check for the difference in terms of students' cognitive presence, we constructed a MANOVA model with cluster assignment as a single independent variable and five dependent variables: four measures of cognitive presence (i.e., the number of messages in four phases of cognitive presence) and the number of non-cognitive messages (coded as other).

Before running MANOVAs, similarly to the work of Lust et al. (2011, 2013a, 2013b), the homogeneity of covariance assumption was checked using Box's M test and homogeneity of variances using Levene's test. To protect from the assumption violations, we log-transformed the data and used the Pillai's trace statistic which is considered to be robust against assumption violations (Field, Miles, & Field, 2012). As the final protection measure, obtained MANOVA results were compared with the results of the robust rank-based variation of the MANOVA analysis using the approach by Nath and Pavur (1985).

In the case of significant MANOVA, a follow-up univariate one-way analyses of variance (ANOVA) were conducted on each of the dependent variables. This use of the univariate follow-ups after a significant multivariate analysis is often considered as a "protection" from the Type I errors arising from the direct use of multiple ANOVAs (Bock, 1985). However, this approach only protects against Type I error inflation for those dependent variables for which a significant multivariate effect was found (Bray & Maxwell, 1985). Thus, in order to further control for the Type-I error rate inflation due to the multiple comparisons, the very conservative Bonferroni correction was adopted. Before running ANOVAs, the homogeneity of variance was checked using Levene's test, and when it was violated, the non-parametric Kruskal–Wallis test was used. Significant Kruskal–Wallis tests were followed up by pairwise comparisons also using the Bonferroni correction. Finally, after significant ANOVAs, Tukey's honest significant difference (HSD) test was used to check for the differences among the individual pairs of clusters.

Given that univariate follow-up analysis does not examine multivariate differences among different conditions, we used the discriminatory factor analysis (DFA), which is very commonly used to assess the multivariate effects of the significant MANOVA analyses (Field et al., 2012). Two approaches together (i.e., ANOVAs and DFA) are considered to provide a complete picture of the multivariate differences among the different groups (Field et al., 2012).

5. Results

5.1. Clustering results

5.1.1. Selecting the number of clusters

Fig. 1 shows the dendogram tree of the student clustering by using the agglomerative hierarchical clustering algorithm. The length of the vertical connecting lines in the dendogram tree indicates the difference between two merged clusters. Starting with the two-cluster solution, more detailed cluster solutions were evaluated, and as a final clustering solution selected the solution with six clusters. Clustering solutions with more than six clusters were only different by having additional clusters with either one or two students, which was indicative that the appropriate number of clusters was selected. Fig. 2 shows the difference between the centers of all the six final clusters, while Table 4 shows the raw scores of clustering variables for each cluster. The systematic relationship between individual course offerings and identified clusters (Fig. 3) was checked, and no clear pattern was observed (Pearson's correlation $r = -0.159$, $p = 0.156$). Finally, to make the reporting of the results easier, each cluster was assigned a label (Table 5) based on our analysis and interpretation of the observed cluster differences. Section 6 provides an in-depth discussion of the clustering results.

5.1.2. Description of identified clusters

In Fig. 2, we can see that students in cluster one (task-focused users) are mostly below the average mean value for all the clustering variables, except those that were related to posting discussion messages. With respect to cluster two (content-focused no-users), it was around average for most variables, except for those variables related to online discussions. The mean values of cluster three (no-users) are below the mean values for all clustering variables, except for the number of logins into the LMS which was around the average. In a complete contrast to cluster three (no-users), the students from cluster four (highly intensive users) had all of their mean values above the overall mean with some of them being several standard deviations larger than the average (e.g., the number of logins into the system and number of course, assignment, and discussion views). Students in cluster five (content-focused intensive users) show more moderate values, with a focus on "non-social" aspects of LMS, while students in cluster six (socially-focused intensive

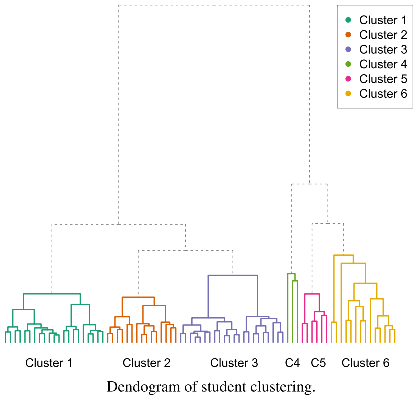


Fig. 1. Dendogram of student clustering.

users) show the opposite trend, primarily focusing on online discussion participation.

5.2. Analysis of cluster differences

In order to check for the statistical significance between the discovered clusters, a one-way multivariate analysis of variance (MANOVA) is conducted with the students' cluster assignment as the single independent variable and the thirteen technology use measures as the dependent variables. As MANOVA requires more data points in each group than the number of dependent variables (Tabachnick & Fidell, 2007), clusters four (highly intensive users) and five (content-focused intensive users) were removed from the analysis, given that they have only three and six students, respectively. The assumption of homogeneity of covariances was tested using Box's M test (Field et al., 2012) which was not accepted. Thus, Pillai's trace statistic was used, as it is more robust to the assumption violations (Field et al., 2012) together with the Bonferroni correction method. A statistically significant MANOVA effect was obtained, Pillai's Trace = 1.62, $F(39, 174) = 5.28$, $p < 10^{-14}$. The multivariate effect size was estimated at multivariate $\eta^2 = .54$, which implies that 54% of the variance in the canonically derived dependent

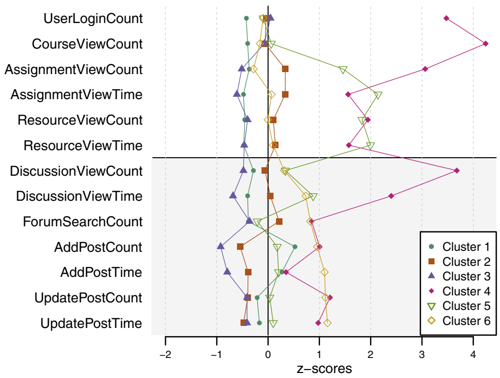


Fig. 2. Clustering results.

variable was accounted for by the differences in the student cluster assignment, which is according to Cohen (1988) and Miles (2001) considered a large effect size. Finally, our findings were further confirmed with robust rank-based MANOVA for which the significant results were also obtained (Wilks $\Lambda_{rank} = 0.06$, $p < 10^{-15}$).

As a follow-up, a series of one-way ANOVAs with Bonferroni corrections for each of the dependent variables was conducted. The assumption of homogeneity of variance was tested using Levene's F test and it revealed that assumption of homogeneity of variance was satisfied for all but two variables (UpdatePostTime and ForumSearchCount) for which the Kruskal–Wallis tests were conducted. All but three ANOVA models (for UserLoginCount, CourseViewCount and ResourceViewCount) were statistically significant (Table 6a), as well as both Kruskal–Wallis tests (Table 6b). The obtained effect sizes were in the range of $\eta^2 = 0.19$ for the number of assignment views to $\eta^2 = 0.62$ for the total number of posted messages. All these are all considered large effect sizes (Cohen, 1988; Miles, 2001). Significant ANOVA models were followed by Tukey's HSD analysis for the pairwise comparison between clusters (Table 6c), and significant Kruskal–Wallis tests were followed by a pairwise Kruskal–Wallis test with the Bonferroni correction (Table 6b).

In addition to the univariate analyses, a discriminatory factor analysis (DFA) was conducted to check for the multivariate differences among the clusters in terms of students' technology use. As the four clusters were selected for the analysis, the DFA produced three discriminant functions (LD1–3) whose standardized loadings are shown in Table 7 and Fig. 4b. As Fig. 4a shows, four clusters are reasonably well separated based on the first two discriminant functions, which accounted for 69% and 22% of the variability in students' cluster assignments, respectively. Coefficients for the first discriminant function (LD1) are mostly positive, with only negative coefficients for the number of course and resource views. In terms of LD1, clusters three (no-users) and six (socially-focused intensive users) represented two extremes, while clusters one (task-focused users) and two (content-focused no-users) were in the middle. On the other hand, coefficients for the second discriminant function (LD2) were mostly negative, except for the positive coefficients for the total time spent on assignments, posting messages and updating messages. Here an opposite trend was observed, with students from clusters one (task-focused users) and two (content-focused no-users) being the extremes, and students in clusters three (no-users) and six (socially-focused intensive users) being in the middle.

5.3. Analysis of cognitive presence

In order to check for the cluster differences in the levels of cognitive presence, a one-way multivariate analysis of variance (MANOVA) was conducted. The independent variable (IV) was the students' cluster assignment and dependent variables (DVs) were the counts of messages in the four different phases of cognitive presence (i.e., triggering event, exploration, integration, and resolution) and the count of messages without traces of cognitive presence (coded as "other"). The descriptive statistics for each of the dependent variables are shown in Table 8.

Before running the MANOVA, cluster four (highly intensive users) with only three students was removed from our analysis, as the MANOVA requires that each condition has more subjects than the dependent variables (Tabachnick & Fidell, 2007) — in this case more than five students as there are five measures of cognitive presence. After the log-transformation of the data, Box's M test for the homogeneity of covariances using the suggested significance level $\alpha = 0.001$ (Tabachnick & Fidell, 2007) indicated that covariance matrices were not significantly different; Box's M = 114.1, $p = 0.005$.

A one-way MANOVA was performed to test the difference between the clusters with respect to the number of messages in the four phases of cognitive presence, as well as the number of non-cognitive messages.

Table 4
Descriptive statistics of cluster centers (raw scores).

		Cluster 1 (N = 21)		Cluster 2 (N = 15)		Cluster 3 (N = 22)		Cluster 4 (N = 3)		Cluster 5 (N = 6)		Cluster 6 (N = 14)	
#	Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	`UserLoginCount`	285	129	420	169	450	297	1650	938	401	101	399	183
2	`CourseViewCount`	403	176	598	242	592	343	3080	1030	668	182	541	294
3	`AssignmentViewCount`	63.9	21.5	87.1	15	59	16.1	177	32.7	124	20.8	66.6	25
4	`AssignmentViewTime`	6.42	2.55	11.5	4.13	5.63	3.4	19.2	6.81	22.8	6.95	9.85	4.27
5	`ResourceViewCount`	31.6	17	50.8	24.3	33.4	15.4	115	53.6	111	48.4	47.1	30.6
6	`ResourceViewTime`	4.07	2.37	8	4.26	4.13	2.46	17.1	12.6	19.8	8.31	7.72	5.64
7	`DiscussionViewCount`	166	39.2	216	91.1	123	57.5	1050	706	304	105	300	134
8	`DiscussionViewTime`	10.8	4.56	15.8	7.81	7.68	5.58	42	15.2	25.1	9.74	23.5	11.5
9	`ForumSearchCount`	0	0	1.13	1.46	0.0455	0.213	2.33	2.08	0.333	0.816	2.29	3.69
10	`AddPostCount`	33.1	6.57	20.9	5.51	16.5	6.14	38.7	18.5	29.2	10	38.1	10.1
11	`AddPostTime`	8.17	3.62	5.34	2.58	3.6	2.04	8.51	5.29	7.89	3.9	11.7	4.33
12	`UpdatePostCount`	5.43	6.3	2.67	3.98	2.45	2.32	27	29.7	9.17	8.75	25.6	25.5
13	`UpdatePostTime`	0.164	0.202	0.0517	0.0855	0.0762	0.0824	0.579	0.603	0.261	0.237	0.645	0.593

Time measures are shown in hours.
1: task-focused users, 2: content-focused no-users, 3: no-users, 4: highly intensive users, 5: content-focused intensive users and 6: socially-focused intensive users.

A statistically significant MANOVA effect was obtained, Pillai's Trace = .76, $F(20, 288) = 3.35$, $p < 10^{-5}$. Similar statistically significant results were confirmed by using a robust rank-based MANOVA, Wilks $\Lambda_{rank} = 0.34$, $p < 10^{-7}$. A multivariate effect size was estimated at multivariate $\eta^2 = .19$, which implies that 19% of the variance in the canonically derived dependent variable was accounted for by the differences in the student cluster assignment, and is considered a large effect size (Cohen, 1988; Miles, 2001).

Before conducting a series of follow-up ANOVAs, the assumption of homogeneity of variance was tested for all five dependent measures using Levene's F test, which was found not significant at $p = 0.05$. A statistically significant difference among student clusters was observed in terms of the number of exploration, integration, and non-cognitive messages, and marginal significance for the number of triggering events (Table 9a). The multivariate η^2 effect sizes ranged from 0.27 to 0.32 which are considered large effect sizes (Cohen, 1988; Miles, 2001). Following the significant results of the ANOVA analyses, a series of post-hoc analyses using Tukey's HSD test was performed and Table 9b shows the pairs of clusters where statistically significant differences were observed. In terms of the number of exploration and non-cognitive messages, the students from clusters one (task-focused users) and six (socially-focused intensive users) had a significantly higher number of messages posted than the students from clusters two (content-focused no-users) and three (no-users). Finally, with respect to the number of integration messages, students from cluster three had a significantly fewer integration messages posted than the students from clusters one (task-focused users), two (content-focused no-users), and six (socially-focused intensive users).

A discriminatory factor analysis (DFA) was also performed to assess the multivariate differences between the student clusters. Given that our independent variable was associated with the five levels (due to the elimination of cluster four (highly intensive users) from the analysis), four discriminant functions were discovered to account for 90%, 8.6%, 0.8% and 0.5% of the variation in the students' cluster assignments, respectively. Their coefficients are shown in Table 10 and in Fig. 5b. The coefficients of the first discriminant function (LD1) indicate that LD1 affected all five dependent variables in the same way, with a focus on integration and exploration messages. However, the second discriminant function (LD2) affected integration and exploration messages in the opposite direction than triggering event, resolution and non-cognitive messages. The coefficients for integration, triggering event, and non-cognitive messages are much bigger than those for the exploration and resolution messages, indicating their much stronger significance for the LD2 scores.

The students' scores in the first two discriminant functions in Fig. 5a were less separated compared to the DFA analysis of the technology use reported in Fig. 4a. The scores in the first discriminant function (Fig. 5c) show that the students from clusters one (task-focused users) and six (socially-focused intensive users) had similar scores; likewise, the students from clusters two (content-focused no-users) and three (no-users) with the students from cluster three having somewhat lower scores. Interestingly, students from cluster five (content-focused intensive users) had very disperse scores for the first discriminant function. The second discriminant function (Fig. 5d) reveals mixed scores, with the students from cluster five (content-focused intensive users) and two (content-focused no-users) in general having somewhat lower and higher scores, respectively.

6. Discussion

6.1. Research question 1: technology-use profiles within communities of inquiry

Based on the results of clustering, we can confirm the existence of the different technology-use profiles within the communities of inquiry. Generally speaking, our findings are aligned with the existing research of students' technology use, with some interesting differences which are discussed in the reminder of this section. What is particularly interesting are the magnitudes of the obtained effect sizes. Both MANOVA

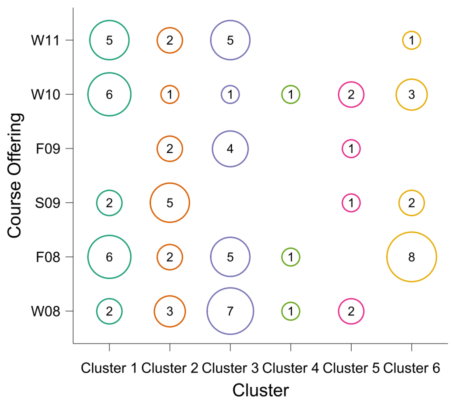


Fig. 3. Distribution of clusters across course offerings.

Table 5
Summary of cluster differences.

#	Size	Label	Characteristics
1	21	Task-focused users	Overall below average activity, Above average message posting activity
2	15	Content-focused no-users	Below average discussions-related activity, Average content-related activity, emphasis on assignments
3	22	No-users	Overall below average activity, slightly bigger in discussion-related activities
4	3	Highly intensive users	Significantly most active students, especially in content-related activities
5	6	Content-focused intensive users	Above average content-related activity, Average discussion-related activity
6	14	Socially-focused intensive users	Above average discussion-related activity, Average content-related activity

and the subsequent ANOVA effect sizes are all very large, suggesting important differences between students in terms of their technology use. This is especially evident for discussion-related activities, for which the obtained effect sizes are particularly large. This suggests stronger differences between students' technology-use profiles with respect to the use of asynchronous online discussions than to the use of the course content.

The DFA results (Fig. 4) show that the first linear discriminant (LD1) – which accounts for almost 70% of the variability in students' cluster assignments – could be best described as *the amount of the overall engagement*, with the focus on the quality of discussion-related activity. This is aligned with the previous work of Lust et al. (2011, 2013a, 2013b) and Valle and Duffy (2009) who found large differences between students in terms of the effort invested in the course. The differences in engagement were also expected, given the different conditions on which students regulate their own learning activities (Butler & Winne, 1995; Winne, 2006; Winne & Hadwin, 1998). The students from cluster six (socially-focused intensive users) have the highest LD1 scores, while the students from cluster three (no-users) have the lowest scores. The students from cluster one (task-focused users) have medium scores, primarily due to their high message posting activity, while the students from cluster two (content-focused no-users) also have medium scores, mostly due to their intermediate content-related activities. Likewise, the high cost of the last merging step on the clustering dendogram (Fig. 1) also points out on the substantial differences between the last two clusters. The observed differences are also consistent across all the variables suggesting that *the overall engagement with the system is the most important characteristic that defines the students' technology-use profile*.

The second discriminant function (LD2) – which accounted for 22% of the variability in the student cluster assignment – could be best described as *the focus on message posting activity or the preference towards discussion participation*. This is aligned with the findings of Wise et al. (2013) who identified three groups of students with different preferences towards active vs. passive participation in online discussions. The study of Dennen (2008) found that students' engagement in asynchronous online discussions was related to their perceived usefulness of learning through discussions, which is aligned with the notion of students' self-regulation of learning activities (Winne & Hadwin, 1998). Likewise, Bliuc et al. (2010) showed that students' conception of learning through discussions (i.e., cohesive vs. fragmented) was related to their approaches to learning and ultimately academic performance.

In the study reported in this paper, the students from cluster one (task-focused users) have the highest scores of LD2 due to their

Table 6
Cluster comparison results.
(a) ANOVA results. Boldface indicates statistical significance at level $\alpha = 0.0038$ (0.05/13).

Variable	Levene's		ANOVAs		
	$F(3, 68)$	p	$F(3, 68)$	p	η^2
`UserLoginCount`	2.23	.09	2.05	.12	.08
`CourseViewCount`	0.77	.51	2.04	.12	.08
`AssignmentViewTime`	**1.62**	**.19**	**8.18**	**.0001**	**.27**
`AssignmentViewCount`	**2.41**	**.07**	**5.23**	**.003**	**.19**
`ResourceViewTime`	0.31	.82	4.29	.008	.16
`ResourceViewCount`	0.69	.56	2.09	.11	.08
`DiscussionViewTime`	**1.02**	**.39**	**13.98**	**<.0001**	**.38**
`DiscussionViewCount`	**2.52**	**.07**	**19.41**	**<.0001**	**.46**
`AddPostTime`	**0.27**	**.85**	**16.77**	**<.0001**	**.43**
`AddPostCount`	**2.71**	**.052**	**38.41**	**<.0001**	.62
`UpdatePostCount`	1.06	.37	14.15	**<.0001**	**.38**

(b) Kruskal–Wallis results and posthoc analysis results.

Variable	$H(3)$	p	Cluster pair
`ForumSearchCount`	**22.38**	$<\mathbf{10^{-4}}$	6–1
			6–3
			2–1
			2–3
`UpdatePostTime`	**24.16**	$<\mathbf{10^{-4}}$	6–1
			6–2
			6–3

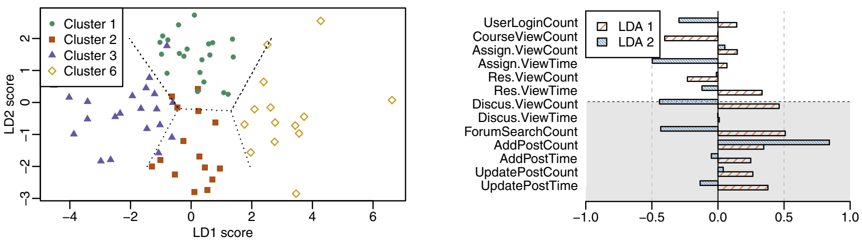
(c) ANOVA posthoc analysis results.

Variable	Cluster pair	Difference	p adjusted
`AssignmentViewTime`	2–1	0.755	0.012
	2–3	1.033	0
	6–3	0.825	0.006
`AssignmentViewCount`	2–1	0.504	0.013
	2–3	0.597	0.002
	2–6	0.464	0.049
`DiscussionViewTime`	6–1	1.050	0.001
	2–3	1.094	0
	6–3	1.628	0
`DiscussionViewCount`	1–3	0.538	0.007
	6–1	0.783	0
	2–3	0.855	0
	6–3	1.321	0
`AddPostTime`	1–3	1.000	0
	6–2	1.059	0
	6–3	1.501	0
`AddPostCount`	1–2	0.661	0
	1–3	1.024	0
	2–3	0.363	0.036
	6–2	0.841	0
	6–3	1.204	0
`UpdatePostCount`	6–1	2.107	0
	6–2	2.768	0
	6–3	2.692	0

1: Task-focused users, 2: content-focused no-users, 3: no-users, 4: highly intensive users, 5: content-focused intensive users, 6: socially-focused intensive users.

Table 7
Standardized coefficients of discriminant functions.

Variable	LD1	LD2	LD3
`UserLoginCount`	0.14	−0.3	0.24
`CourseViewCount`	−0.4	−0.001	0.14
`AssignmentViewCount`	0.15	0.052	−0.74
`AssignmentViewTime`	0.068	−0.5	0.08
`ResourceViewCount`	−0.23	−0.012	0.48
`ResourceViewTime`	0.33	−0.12	−0.42
`DiscussionViewCount`	0.46	−0.44	−0.26
`DiscussionViewTime`	0.01	−0.002	0.07
`ForumSearchCount`	0.51	−0.43	−0.019
`AddPostCount`	0.35	0.84	−0.44
`AddPostTime`	0.25	−0.05	0.28
`UpdatePostCount`	0.26	0.04	0.32
`UpdatePostTime`	0.38	−0.14	0.44
`Variance Explained`	0.69	0.22	0.091



(a) Scores for the first two discriminant functions. The DFA decision boundaries are shown with the dotted lines.

(b) Loadings of the clustering variables on the first two discriminant functions.

Fig. 4. Results of the discriminant function analysis for the multivariate differences between clusters in terms of the technology use. 1: Task-focused users, 2: content-focused no-users, 3: no-users, 4: highly intensive users, 5: content-focused intensive users, 6: socially-focused intensive users.

unusually high posting activity and the overall low engagement with the LMS. In contrast, the students from cluster two (content-focused no-users) – given their average engagement with the LMS – have unusually low message posting activity. This is also evident from the cluster differences (Fig. 2) which indicate that different preferences guide students' choices when exercising their agency (Winne, 2006), in this case, towards online discussions. The students from cluster two (content-focused no-users), three (no-users), four (highly intensive users), and five (content-focused intensive users) clearly exhibit strong inclination towards content-related activities (assignments and resources) and passive reading of the discussion messages, while the students in cluster one (task-focused users) and six (socially-focused intensive users) exhibit higher inclination towards participation in asynchronous online discussions.

In order to summarize our findings and to make them more usable, Fig. 6 integrates all the previously described results into a coherent picture of cluster differences using two dimensions: i) the level of content-related activity (including discussion reading), and ii) the level of active participation in online discussions. We can see that students adopted different study tactics which might be related to the differences in their internal and external conditions, particularly metacognition and motivation as defined by the previous research (Lust et al., 2013b).

Previous research suggests the existence of three or four technology-use profiles, while our results indicate the existence of six different profiles. Indeed, Lust et al. (2011) points out on the possible existence of additional clusters that are harder to detect empirically due to the diminishing differences between them. On the other hand, it is possible that the larger number of clusters identified in this study is just a reflection of anomalies and outliers in the data, particularly due to existence of some smaller clusters such as clusters four (highly intensive users) and five (content-focused intensive users). Concerning the relative sizes of the clusters, previous studies by Lust et al. (2011, 2013a, 2013b) found that no-users are the largest group, and that intensive users are the smallest. This is consistent with the sizes of the clusters in our study, as cluster three (no-users) and four (highly intensive users) corresponded reasonably well with no-users and intensive users, respectively. Regarding the previously discovered selective user cluster (Lust et al., 2011, 2013a, 2013b), our study found several user profiles that might be called selective. One possible reason might be the fact that the course in our study was a graduate level course with many of the students having previous work experience and completed bachelor's degrees. It is shown that more experienced students who had previous education and work experience were able to make better use of the discussion boards (Lust et al., 2011). Likewise, given that the course in our study was from a fully online program, the students were already familiar with this type of learning environments, and thus better able to manage their own learning processes.

An interesting finding by Lust et al. (2011) is that the intensive cluster was the most diverse in terms of their technology use. Our study confirms these findings. The clustering dendogram shows (Fig. 1) that clusters one (task-focused users), two (content-focused no-users), three (no-users), and five (content-focused intensive users) became fully connected while the three students from cluster four (highly intensive users) remained in isolated single-element clusters. As the students in cluster four exhibited more diversified behavior, it resulted in the larger differences in their scores on the used clustering variables, which subsequently resulted in their later merging into a single cluster.

6.2. Research question 2: effects of technology use on cognitive presence

The results of the MANOVA analysis indicate that there is a significant difference between the students in different clusters in terms of their cognitive presence. We find a large effect size, as the cluster assignment accounted for 19% of the variability in the canonically derived dependent variable, which suggests an important connection between technology use and the students' cognitive presence. Furthermore, the results of the subsequent ANOVA analyses suggest that differences were strongest for the integration phase, followed by the differences for the exploration phase and with the smallest differences for the number of non-cognitive messages. The differences in the number of

Table 8
Descriptive statistics of the dependent variable raw scores: median (Mdn), 25th (Q1) and 75th (Q3) percentiles.

		Cluster 1 (N = 21)		Cluster 2 (N 15)		Cluster 3 (N = 22)		Cluster 4 (N = 3)		Cluster 5 (N = 6)		Cluster 6 (N = 14)	
#	Variable	*Mdn*	(*Q1*, *Q3*)	*Mdn*	(*Q1*, *Q3*)	*Mdn*	(*Q1*, *Q3*)	*Mdn*	(*Q1*, *Q3*)	*Mdn*	(*Q1*, *Q3*)	*Mdn*	(*Q1*, *Q3*)
1	`TriggeringEventCount`	3	(1, 5)	2	(1, 2)	2	(1, 3.8)	2	(1.5, 9.5)	3.5	(2, 6.5)	4.5	(2.2, 5)
2	`ExplorationCount`	10	(6, 11)	5	(3.5, 8)	4.5	(3, 6)	10	(6.5, 12)	9	(4.8, 12)	11.5	(8.5, 18)
3	`IntegrationCount`	8	(4, 11)	6	(4.5, 8)	3	(2, 3)	6	(6, 9.5)	4	(4, 8.5)	8	(5.2, 13)
4	`ResolutionCount`	1	(0, 2)	1	(0, 1.5)	0	(0, 1)	1	(0.5, 2)	1	(0.25, 2.5)	1.5	(0, 2.8)
5	`OtherCount`	9	(5, 11)	4	(3, 6)	4.5	(3.2, 7.5)	15	(10, 16)	9	(9, 9.8)	9.5	(5.2, 14)

1: Task-focused users, 2: content-focused no-users, 3: no-users, 4: highly intensive users, 5: content-focused intensive users, 6: socially-focused intensive users.

Table 9
Cognitive presence analysis results.

(a) ANOVA Results. Boldface indicates statistical significance at level $\alpha = 0.01(0.05/5)$.

Variable	Levene's		ANOVAs		
	$F(4.73)$	p	$F(4.73)$	p	η^2
`Trig.EventCount`	1.85	.13	2.69	0.038	.13
`ExplorationCount`	**1.14**	**.34**	**7.71**	**<.0001**	**.30**
`IntegrationCount`	**0.66**	**.62**	**8.88**	**<.00001**	**.32**
`ResolutionCount`	1.09	.37	1.57	.19	.08
`OtherCount`	**1.00**	**.41**	**6.79**	**<.001**	**.27**

(b) Significant pairwise comparisons of cluster centers.

Variable	Cluster Pair	Difference	P adjusted
`ExplorationCount`	1–2	0.745	0.029
	1–3	0.889	0.002
	6–2	1.012	0.004
`IntegrationCount`	6–3	1.156	0
	1–3	1.194	0
	2–3	0.959	0.004
	6–3	1.288	0
`OtherCount`	1–2	0.843	0.007
	1–3	0.763	0.007
	6–2	0.940	0.006
	6–3	0.861	0.007

1: Task-focused users, 2: content-focused no-users, 3: no-users, 4: highly intensive users, 5: content-focused intensive users, 6: socially-focused intensive users.

triggering event messages are marginally significant, which can be explained by the grading policy for discussions and their external motivation induced by the course design — which is very common in higher education (Garrison et al., 2001; Penny & Murphy, 2009; Rovai, 2007). Likewise, the lack of significance for the resolution phase is not surprising as – due to the time constraints – in most higher education settings the resolution phase is not reached (Garrison et al., 2001; Garrison, Anderson, et al., 2010a).

The DFA results (Fig. 5) reveal that the first discriminant function (LD1) – which accounts for 90% of the variability between students' cluster assignments – can be summarized as *the development of cognitive presence*. LDA1 makes a strong distinction between students from clusters one (task-focused users), five (content-focused intensive users), and six (socially-focused intensive users) on the one side, and students from clusters two (content-focused no-users) and three (no-users) on the other side. The distribution of scores within clusters is reasonably consistent, with only students from cluster five (content-focused intensive users) scoring more diversified scores. The observed differences were expected based on the previous research (Clarebout et al., 2013; Lust et al., 2011, 2013a, 2013b; Valle & Duffy, 2009; Wise et al., 2013), and it was not surprising that the students from less engaged clusters do not fully develop their cognitive presence. It is aligned with existing evidence of students' poor self-regulation of learning (Dunlosky & Lipko, 2007). Likewise, there is evidence suggesting that a majority of the students do not develop their cognitive presence past the exploration phase (Garrison et al., 1999).

The ANOVA results (Table 9) indicate that the students in clusters two (content-focused no-users) and three (no-users) have significantly fewer exploration and non-cognitive messages than students from

Table 10
Standardized coefficients of discriminant functions.

Variable	LD1	LD2	LD3	LD4
`TriggeringEventCount`	0.1	0.44	0.57	−0.87
`ExplorationCount`	0.72	−0.13	−0.9	0.11
`IntegrationCount`	0.67	−0.65	0.21	0.013
`ResolutionCount`	0.16	0.034	0.35	−0.39
`OtherCount`	0.49	0.62	0.3	0.54
`VarianceExplained`	0.9	0.086	0.008	0.005

clusters one (task-focused users) and six (socially-focused intensive users), and that students in cluster three (no-users) have significantly fewer integration messages than students from clusters one (task-focused users), two (content-focused no-users), and six (socially-focused intensive users). Based on this, it is evident that the lack of cognitive presence development is more emphasized for the students in cluster three (no-users), which is aligned with their overall low level of the use of the LMS.

The second discriminant function (LD2) (Fig. 5b and Table 10) – which explains 9% of the variability of students' cluster assignment – can be best summarized as the relative *lack of integration*, given the overall level of development of cognitive presence. In general, LD2 scores of students from all the clusters are far less separated. This find is reasonable given the smaller amount of variance explained by LD2. Still, it can be seen that students from cluster five (content-focused intensive users) had slightly lower LD2 scores than other students, and that cluster one (task-focused users) and two (content-focused no-users) students had slightly higher LD2 scores. Given their high engagement and clear preference towards content-related activities, one likely explanation could be their lower perceived usefulness of learning through discussions (Dennen, 2008) and more fragmented conception of learning through discussions (Bliuc et al., 2010).

What is very interesting is that *we do not observe a clear connection between the overall engagement and the development of cognitive presence*. Although students from clusters two (content-focused no-users) and three (no-users) have a lower level of the use of the LMS and had lower levels of cognitive presence, this is not the case for the students in cluster one (task-focused users). Similar results are already found in the literature. The study by Valle and Duffy (2009), concluded that adoption of different learning profiles have very little impact on the final course grade, with even the students who struggled with online environment finishing the course with similar final grades. Likewise, the study by Wise et al. (2013) found no difference in terms of the final grades between students with different profiles of participation in online discussions. Given that the course under investigation in our study is a graduate level course, it is likely that students in our study posses higher meta-cognitive awareness and motivation (Perkins, 1985; Winne, 2006; Winne & Hadwin, 1998), as well as broader domain knowledge and skills (Lust et al., 2011).

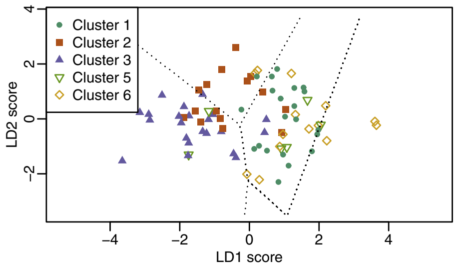
6.3. Cluster interpretations

This section provides more detailed interpretations of the identified clusters by building on the existing research on the CoI model and learning technology use and, and important constructs that are identified in the previous studies (e.g., motivation, self-efficacy, self-regulated learning, and goal orientation). The implications that can inform future research and practice are also discussed.

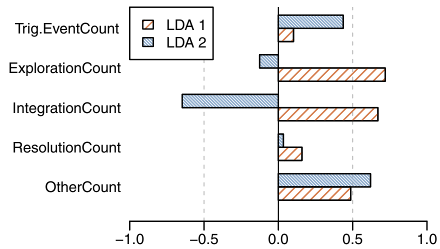
6.3.1. Cluster one (task-focused users)

The students from cluster one show below average use of the LMS, except for the number of posted messages and time spent on writing new messages. Even though they have above average message posting activity, their reading activity is below average. They resemble very closely task focused, "get it done" students from the Valle and Duffy (2009) study and they tend to spend only the time necessary for completion of the course. As Valle and Duffy explain, they seem to be in a hurry to complete the course, but still show positive study strategies comparable to mastery-oriented and self-driven students. Given that a large number of the students in our study are also working full time, this might be one of the reasons for the observed behavior of this cluster of students.

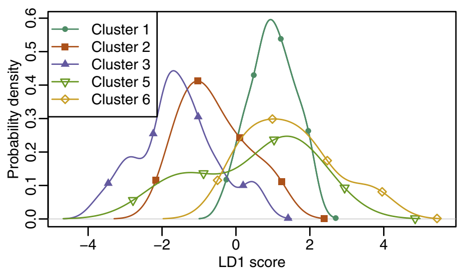
Regardless of the LMS use, the levels of cognitive presence are very high for the students in this cluster. This further supports findings by Valle and Duffy (2009) of the use of positive study strategies. The students from cluster one have significantly more triggering events and



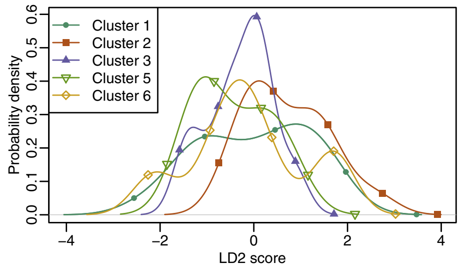
(a) Scores for the first two discriminant functions. LDA decision boundaries are shown with dotted lines.



(b) Loadings of clustering variables on the first two discriminant functions.



(c) Distribution of scores for the first discriminant function (LD1)



(d) Distribution of scores for the second discriminant function (LD2)

Fig. 5. Results of the discriminant function analysis for the multivariate differences between clusters in the levels of cognitive presence. 1: Task-focused users, 2: content-focused no-users, 3: no-users, 4: highly intensive users, 5: content-focused intensive users, 6: socially-focused intensive users.

non-cognitive messages than those of the students from clusters two (content-focused no-users) and three (no-users), and significantly more integration messages than those of the students from cluster three (no-users). They were very similar to the students from cluster six (socially-focused intensive users), while being significantly less active overall. Similar results are reported by Lust et al. (2011) who did not find significant differences between intensive and incoherent users in terms of their academic performance. Given that quality rather than quantity of the LMS use is indicative of students' metacognitive skills (Clarebout et al., 2013) (also consistent with the learning strategy use as suggested by Kuhn (1995)), it seems likely that the students in cluster one possess higher metacognitive skills that enable them to control and monitor their learning activities effectively. As pointed out by Lust et al., the differences in tool-use patterns are not necessarily a problem, especially in learning environments which focus on active, self-controlled learning. However, if its affects the students' performance, it might suggest that those students are not profiting from the whole range of tools that are available to them (Lust et al., 2011).

Looking at the cluster sizes (Fig. 1) we can see that cluster one is relatively large (N = 21) – suggesting that a considerable number of students were meta-cognitively skillful enough in the use of asynchronous online discussions. This is aligned with the results of Valle and Duffy who also looked at effectiveness of students in online learning settings. A possible explanation is that their previous experience with this particular learning modality – given that the students were enrolled in a fully online condition – enables them to be more effective with their learning. This also explains why studies that looked at metacognitive skills of students in blended learning courses (Lust et al., 2013a) showed that a majority of students did not adequately regulate their tool use. Still, this interpretation warrants more empirical and theoretical attention in future studies.

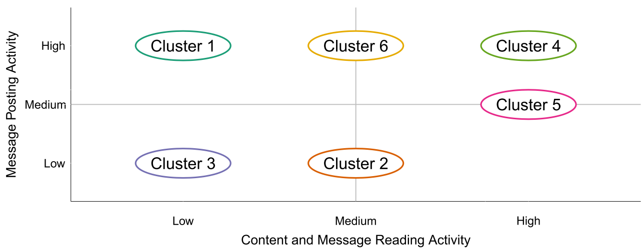


Fig. 6. Cluster matrix: activity focus and activity level. 1: Task-focused users, 2: content-focused no-users, 3: no-users, 4: highly intensive users, 5: content-focused intensive users, 6: socially-focused intensive users.

The fact that the students from cluster one spend very little time reading other students' writings is very interesting. This might suggest that there was much of "scanning activity" by the students. That type of activity would unlikely result in deep comprehension (Hewitt, Brett, & Peters, 2007). As indicated by Entwistle (2009), effective learning can be combined with both deep and surface approaches to learning. Also, students who post solely to meet the course requirements and read the bare minimum of other students' writings are less likely to perceive online discussions as a valuable learning activity (Dennen, 2008). However, those types of students benefit particularly by requesting the student participation in asynchronous online discussions in the course design (Wise et al., 2013). This might be an important reason for their high levels of cognitive presence, which is empirically shown to be shaped by teaching presence (Garrison, Cleveland-Innes, et al., 2010).

6.3.2. Cluster two (content-focused no-users)

The students from cluster two are characterized by the overall average LMS use. They only show above average scores for the assignment-related variables. With respect to discussion-related activities, they spend about an average amount of time reading discussions, but seem reluctant to contribute themselves. They are most similar to the selective and incoherent users from the Lust et al. (2011, 2013b) studies, respectively.

The average time spent reading discussions and the below-average time spent actively participating suggests fragmented conceptions of learning through discussions (Bliuc et al., 2010) and the lack of metacognitive knowledge and skills required for successful participation in online discussions (Hew, Cheung, & Ng, 2010). The study by Lust et al. (2013b) indicated that metacognitive skills were mostly influencing students' tool use patterns through goal orientation. Given the focus of the students from cluster two on assignments and their similarity to selective users of Lust et al. (2013b), it seems likely that the students in cluster two have high inclination towards performance goal orientation, which is in turn shown to be related to surface approaches to learning (Phan, 2008).

With respect to the level of cognitive presence, the students from cluster two show generally lower levels of cognitive presence. The ANOVA results (Table 9) show that they have a significantly lower number of integration and non-cognitive messages than that of the students from clusters one (task-focused users) and six (socially-focused intensive users). This is also confirmed by the DFA results (Fig. 5) given the low scores of students from this cluster on the first discriminant function (LD1) (Fig. 5c). Furthermore, the high scores of cluster two on the second discriminant function (LD2) (Fig. 5d) indicates a lower proportion of integration messages relative to the number of triggering event and non-cognitive messages. This offers further evidence of the lack of skills for successful participation in online discussions and their need for better instructional support in order to overcome challenges that are associated with participation in discussions and inquiry-based learning (Cho & Kim, 2013).

6.3.3. Cluster three (no-users)

With 22 students, cluster three is the largest in size. The primary characteristics of students from this cluster is their below-average overall engagement with the LMS, as their scores for all variables (except for number of logins) are below the overall mean. Furthermore, ANOVA results indicate that for many of the variables, they have significantly lower scores than students from other clusters (Table 6). In addition, they have the lowest scores for the first discriminant function (Fig. 4) which confirms their low use of the LMS. Looking at the findings from the previous studies, the students from this cluster are most similar to the no-users from the Lust et al. (2011, 2013a, 2013b) studies and procrastinating, "minimalist in effort" students from the Valle and Duffy (2009) study. The low use of the LMS is also reflected in their lower levels of cognitive presence. The students from cluster three have significantly less exploration, integration, and non-cognitive messages than students from clusters one (task-focused users) and six (socially-focused intensive users), as well as a significantly lower number of integration messages than those of the students from cluster two (content-focused no-users). The size of this cluster and the limited cognitive presence development is aligned with the previous research by Garrison et al. (2001).

There are several possible explanations for the observed characteristics of students in cluster three. One likely cause might be the lack of intrinsic motivation to engage in learning as – according to Valle and Duffy (2009) – the way in which students learn online might indicate their commitment to learning. Much like students from cluster two (content-focused no-users), it is possible that students in this cluster have high levels of performance goal orientation and surface approaches to learning (Phan, 2008). Finally, it is interesting to note that the low use of the LMS is especially evident with respect to discussion-related activities. This might suggest that the students from cluster three lack the metacognitive skills required for successful learning in this particular context (Valle & Duffy, 2009). Drawing on the results from Bliuc et al. (2010), it is also very likely that they follow fragmented approaches to learning in online discussions, which is also aligned with the research indicating weakness in regulation of learning by a large proportion of students (Dunlosky & Lipko, 2007; Lust et al., 2013a). As suggested by Cho and Kim (2013), the students three may also require better instructional support for interaction with others — similar to the students from cluster two (content-focused no-users).

6.3.4. Cluster four (highly intensive users)

With only three students, cluster four is the smallest in our study. Although it was excluded from the statistical analyses, it is clear that students from this cluster are characterized by a very high level of motivation and the use of the system (Fig. 2). Students from this cluster log into the LMS much more often than students from other clusters, and they also frequently check the state of the discussions and assignments. Looking at Fig. 2 and Table 4, it is clear that the most of their count measures are very high, some even several standard deviations above the overall mean (e.g., the number of course logins, assignment views, and discussion views). Due to their overall high use of the LMS, they show similarities to the intensive active students from the Lust et al. (2013b) study, and the mastery oriented, "self-driven" students from the Valle and Duffy (2009) study. The results of both the Lust et al. (2013b) and Valle and Duffy (2009) studies suggest that mastery goal orientation is associated with this cluster of students. In particular, Lust et al. (2013b) suggest the association with mastery-avoidance goal orientation as it can explain the overly high use of the LMS system.

The students in cluster four are also characterized by the largest number of discussion views, 6.3 and 3.5 times as many as those of the students from clusters one (task-focused users) and six (socially-focused intensive users), respectively (Table 4). They also spend most of their time reading discussions, 3.9 and 1.8 times as much as that of the students from clusters one and six, respectively. However, in terms of the average time spent on reading discussions, cluster four students spend substantially less time on average than students from clusters one and six. Students from cluster four spend on average 2.4 min on each discussion reading activity, while students from clusters one and six spend 3.9 and 4.7 min, respectively (Table 4). Therefore, they were mostly similar to the "broad listeners, reflective speakers" cluster from the study of Wise et al. (2013) who showed similar inclination towards broad reading of discussions.

With respect to active participation in the discussions, the students from cluster four write a similar number of messages as that of the students from cluster six (socially-focused intensive users), but spend much less time on writing their responses. The less time spent on writing messages might be caused by: i) posting many non-cognitive messages, which typically require much less effort to write (Joksimović, Gašević, Kovanović, Adesope, & Hatala, 2014), ii) posting slightly fewer integration messages, which require the most time to write, and iii) a longer discussion reading time which is often related to high comprehension and depth of learning (Wise et al., 2013). The number of messages in different phases of cognitive presence are similar to those of students from cluster one and six (Table 8), with a slightly smaller number of the integration messages. The students in cluster four also have the highest number of non-cognitive messages. As Wise et al. (2013) results showed, "broad listeners, reflective speakers" were the most frequent posters who attended to almost all of the peers' messages with the most posts per discussion. Thus, the large number of non-cognitive messages – which are typically acknowledgments of others' contributions – can be explained by their high attendance to postings of other students. Although our data do not warrant drawing conclusive interpretations, it is likely that the intensive discussion reading – coupled with lower numbers of integration messages and large numbers of non-cognitive messages – is due to the high levels of motivation and lower levels of meta-cognitive skills for online discussions. However, this needs to be further investigated in future studies.

This group is the most diverse in terms of their technology use, as indicated by the late merging of this cluster in the dendogram (Fig. 1). This is aligned with the findings of Lust et al. (2011) who also found the highest divergence in the most engaged group. The diversity in technology-use is also shown to be associated with higher meta-cognitive activity (Lust et al., 2013a), as well as to be an indicator of metacognitive-monitoring (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007).

6.3.5. Clusters five (content-focused intensive users)

Cluster five is mainly characterized by the focus on the course content and passive reading of online discussions, much like the students from cluster two (content-focused no-users); however, with the overall much higher use of the LMS. The students from cluster five have around average number of LMS logins and course views, coupled with high number of assignment and resource views. The students from this cluster also spend more time on assignments and static resources than the students from any other cluster. They also spend more time reading discussions than most of the students (i.e., all students except students from cluster four). However, their active participation in the discussions through message posting is just slightly above the overall mean value.

Similarly to cluster two students (content-focused no-users), students from cluster five show similarity with incoherent users from the Lust et al. (2011) study and selective users from the Lust et al. (2013b) study — given by their clear inclination towards the use of static course content and passive reading of online discussions. The Lust et al. (2011) study found no differences between the incoherent users and the intensive users, and our study showed similar results (i.e., posthoc analysis found no difference between cluster five and any other cluster, as shown in Table 9b).

Based on the Lust et al. (2013b) results, the students from cluster five likely have higher levels of performance-approach orientation than other high technology users (i.e., socially-focused intensive users from cluster six). Similarly, they likely have higher levels of mastery-avoidance orientation than students from clusters two (content-focused no-users) and three (no-users). This could explain their focus on static course information and discussion "consumption" rather than active participation. Given their reluctance to participate, it is likely that those students lack skills that are required for successful participation in the discussions (Hew et al., 2010). Drawing on the Bliuc et al. (2010) results, it is likely that they have high levels of fragmented conceptions of learning in asynchronous online discussions. Thus, the cluster five students might need better instructional support and scaffolding in order to successfully participate in asynchronous online discussions (Cho & Kim, 2013).

The median values of the numbers of messages in each phase of cognitive presence (Table 8) indicate that the students from cluster five have a similar number of messages as the students from clusters one (task-focused users) and six (socially-focused intensive users), except for the lower number of integration messages. This is also visible in the DFA results, as the LD1 scores of the students in cluster five are similar to those of the students from clusters one and six, but more disperse. The lower number of integration messages also reflects on their LD2 scores, which are the lowest for the students in this cluster. Still, as Tukey's HSD analysis do not indicate any statistically significant difference from the students from other clusters, we cannot draw conclusive inferences from the observed data and future research to further examine this cluster is warranted.

6.3.6. Cluster six (socially-focused intensive users)

Students from cluster six are characterized by an average content-related activity and above average discussion-related activity. Their scores on all discussion-related variables (except the number of discussion views) are around one standard deviation above the overall mean values — indicating a strong commitment to learning through asynchronous online discussions. By having a limited number of sessions in which they spent a significant amount of time, they show a similarity with "concentrated listeners, integrated talkers" from Wise et al. (2013) which suggest the depth of their reading activities. However, they also show a certain similarity with the selective users from the Lust et al. (2013b) study, as they are clearly inclined towards learning through online discussions. Based on this, it is likely that those students have cohesive conceptions of learning through discussions (Bliuc et al., 2010), as well as mastery goal orientation. It is interesting to note that they have a higher number of forum searches, suggesting a more strategy approach to the use of online discussions.

The development of the cognitive presence of the students from cluster six is characterized by the high levels of cognitive presence, as shown by the ANOVA (Table 9) and DFA (Fig. 5) analyses. They have significantly more exploration, integration, and non-cognitive messages than those of the students from cluster three, as well as higher number of exploration and non-cognitive messages than students from cluster two (content-focused no-users). In addition, their LD1 scores are very high, confirming their overall high development of cognitive presence. As such, those students might be good candidates for student moderators that are given a responsibility to guide discussions in the productive directions and to assist other students in their own learning (Schellens, Keer, Wever, & Valcke, 2007). Given that student-centered discussions are shown to better foster the development of cognitive presence than instructor-centered discussions (Schrire, 2006), this seems as one promising direction for further research that warrants further empirical examination.

6.4. Limitations

The most important limitations of this study are related to its internal validity as – given its correlational nature – the claims about causality are not truly possible as in the case of randomized controlled trials. Likewise, the sample of 81 students – even though it consists of the six offerings of a course – is still small (although on a higher end of the related studies about communities of inquiry that used quantitative content analysis) and could be affecting the validity of our findings. The data originated from a single graduate-level course at a single university, so the external validity of our findings could be potentially compromised by the specifics of the adopted pedagogical approach in the target course. Furthermore, despite the same course design and organization, the variations in cohort sizes between different course offers could also potentially influence the student learning activities by affecting the climate and overall volume of online discussions which in turn will have an impact on the clustering results. With respect to the construct validity, the cognitive presence construct – which is a latent construct by definition – is measured only through content analysis of discussion transcripts. Finally, the calculated durations for different activities are approximations which could be affected by the current limitations in tracking student activities in the LMS systems.

In order to define technology-use profiles, we adopted clustering techniques, as commonly done in the related studies that looked at learning technology-use profiles. However, clustering is an unsupervised machine learning technique and inherently subjective. Given that there are no upfront right and wrong clustering solutions, this leaves a space for subjectivity in the interpretation of the cluster findings which may or may not affect the final outcomes of the study. In order to interpret our clustering solutions, we looked at the existing literature which provided the necessary foundation for cluster interpretation. However, this brings certain challenges, as the reported findings might not be applicable in the context of the study presented in this paper. In future studies, the explicit use of standard instruments for measuring goal orientation, motivation, and other constructs will be used to provide more empirical evidence for cluster interpretations.

7. Conclusions

The analysis of the clustering process and the final clusters reported in this paper showed several interesting implications for both the educational practice and research on the community of inquiry. Aligned with the previous studies (Clarebout et al., 2013; Lust et al., 2012), our results indicate that the availability of different tools in a learning environment is not enough for their successful use. As indicated by Perkins (1985), students need to be sufficiently meta-cognitively capable, skillful, and motivated to use the available tools. The preference towards static content or discussions and different uses of the available tools suggests a need for different instructional interventions and support for different groups of students. Students that are reluctant to participate in online discussions or who have performance goal orientation might require more detailed instructions on how to productively participate in discussions (Hew et al., 2010), or access to various types of contextual aids, such as access to different static educational resources (Azevedo, 2005). As pointed out by Wise et al. (2013, p. 340): "*Students who are oriented toward mastery and see discussions as vehicles to support this goal are likely to participate in productive ways. In contrast, for students oriented toward performance goals, explicitly embedding desirable participation behaviors in the activity requirements and assessment scheme can help encourage more productive listening and speaking*". Other students, such as students from cluster three (no-users) might require more motivational support, which is aligned with the suggestions of Valle and Duffy (2009).

In terms of the cognitive presence development, our results indicate large differences among students in terms of their cognitive presence (multivariate $\eta^2 = 0.19$). However, low use of the LMS is not necessarily indicative of poor cognitive development. Our results suggest that the quality of activity is more important than quantity as highly metacognitively skilled students – such as students from cluster one (task-focused users) – can be equally successful as more engaged students (i.e., clusters four, five and six). This has also been suggested by Valle and Duffy (2009), Lust et al. (2011), and Clarebout et al. (2013) and our results provide further evidence for this.

Finally, our results suggest that hierarchical clustering can be successfully used for understanding the differences between students in online learning contexts. It is likely that this approach could be used in similar studies and to further validate our findings in other contexts. There are several reasons why hierarchical clustering appears appropriate for these types of studies: i) the number of clusters is in general very small which makes the dendogram analysis manageable and practical, ii) the analysis of the order of cluster merges allows for observing similarities between clusters, thus giving us a view at the factors behind the clustering process, and iii) it can be performed well for small datasets, as it does not depend on the statistical properties of the large datasets like other popular algorithms such as K-means or EM (Abbas, 2008; Hastie et al., 2013).

References

Abbas, O.A. (2008, Jul.). Comparisons between data clustering algorithms. *The International Arab Journal of Information Technology*, *5*(3), 320–325.

Akyol, Z., & Garrison, D.R. (2011a). Assessing metacognition in an online community of inquiry. *The Internet and Higher Education*, *14*(3), 183–190.

Akyol, Z., & Garrison, D.R. (2011b). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology*, *42*(2), 233–250.

Akyol, Z., Arbaugh, J.B., Cleveland-Innes, M., Garrison, D.R., Ice, P., Richardson, J.C., et al. (2009). A response to the review of the community of inquiry framework. *Journal of Distance Education*, *23*(2), 123–135.

Anderson, T., & Dron, J. (2010). Three generations of distance education pedagogy. *The International Review of Research in Open and Distance Learning*, *12*(3), 80–97.

Anderson, T., Rourke, L., Garrison, D.R., & Archer, W. (2001). Assessing teaching presence in a computer conferencing context. *Journal of Asynchronous Learning Networks*, *5*, 1–17.

Arbaugh, J., Cleveland-Innes, M., Diaz, S.R., Garrison, D.R., Ice, P., Richardson, J.C., et al. (2008). Developing a community of inquiry instrument: Testing a measure of the community of inquiry framework using a multi-institutional sample. *The Internet and Higher Education*, *11*(3–4), 133–136.

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist*, *40*(4), 199–209.

Bjork, R.A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444.

Bliuc, A. -M., Ellis, R., Goodyear, P., & Piggott, L. (2010). Learning through face-to-face and online discussions: Associations between students' conceptions, approaches and academic performance in political science. *British Journal of Educational Technology*, *41*(3), 512–524.

Bock, R.D. (1985). *multivariate statistical methods in behavioral research*. Scientific Software International.

Bray, J.H., & Maxwell, S.E. (1985). *Multivariate analysis of variance*. SAGE.

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Journal of Educational Technology & Society*, *15*(3), 3–26.

Butler, D.L., & Winne, P.H. (1995, Sep.). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245–281.

Cho, M. -H., & Kim, B.J. (2013, Apr.). Students' self-regulation for interaction with others in online learning environments. *The Internet and Higher Education*, *17*, 69–75.

Clarebout, G., Elen, J., Collazo, N.A.J., Lust, G., & Jiang, L. (2013). Metacognition and the use of tools. In R. Azevedo, & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies. No. 28 in Springer International Handbooks of Education*. (pp. 187–195). New York: Springer.

Cohen, J. (1988). The analysis of variance. *Statistical power analysis for the behavioral sciences* (pp. 273–406). Hillsdale, N.J.: L. Erlbaum Associates.

Dawson, S., Tan, J.P.L., & McWilliam, E. (2011). Measuring creative potential: Using social network analysis to monitor a learners' creative capacity. *Australasian Journal of Educational Technology*, *27*(6), 924–942.

Dennen, V.P. (2008, Jul.). Pedagogical lurking: Student engagement in non-posting discussion behavior. *Computers in Human Behavior*, *24*(4), 1624–1633.

Dewey, J. (1910). *How we think*. D.C. Heath & Company.

Dunlosky, J., & Lipko, A.R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, *16*(4), 228–232.

Ellis, R.A., Marcus, G., & Taylor, R. (2005). Learning through inquiry: Student difficulties with online course-based material. *Journal of Computer Assisted Learning*, *21*(4), 239–252.

Entwistle, N. (2009, Sep.). *Teaching for understanding at university: Deep approaches and distinctive ways of thinking*. Basingstoke, Hampshire; New York: Palgrave Macmillan.

Field, A.P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Los Angeles; London: SAGE.

Garrison, D.R. (2007). Online community of inquiry review: Social, cognitive, and teaching presence issues. *Journal of Asynchronous Learning Networks*, *11*(1), 61–72.

Garrison, D.R. (2011). *E-Learning in the 21st century: A framework for research and practice* (2nd ed.). New York: Routledge.

Garrison, D.R., & Akyol, Z. (2013). Toward the development of a metacognition construct for communities of inquiry. *The Internet and Higher Education*, *17*, 84–89.

Garrison, D.R., & Arbaugh, J. (2007). Researching the community of inquiry framework: Review, issues, and future directions. *The Internet and Higher Education*, *10*(3), 157–172.

Garrison, D.R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2–3), 87–105.

Garrison, D.R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23.

Garrison, D.R., Anderson, T., & Archer, W. (2010a). The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education*, *13*(1–2), 5–9.

Garrison, D.R., & Cleveland-Innes, M. (2005). Facilitating cognitive presence in online learning: Interaction is not enough. *American Journal of Distance Education*, *19*(3), 133–148.

Garrison, R., Cleveland-Innes, M., & Fung, T.S. (2010b). Exploring causal relationships among teaching, cognitive and social presence: Student perceptions of the community of inquiry framework. *The Internet and Higher Education*, *13*(1–2), 31–36.

Gonyea, R.M. (2005). Self-reported data in institutional research: Review and recommendations. *New Directions for Institutional Research*, *2005*(127), 73–89 (URL http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=18442595&site=ehost-live).

Gorsky, P., Caspi, A., Blau, I., Vine, Y., & Billet, A. (2011). Toward a CoI population parameter: The impact of unit (sentence vs. message) on the results of quantitative content analysis. *The International Review of Research in Open and Distance Learning*, *13*(1), 17–37.

Hadwin, A.F., Nesbit, J.C., Jamieson-Noel, D., Code, J., & Winne, P.H. (2007, Dec.). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*(2–3), 107–124 (URL http://link.springer.com/article/10.1007/s11409-007-9016-7).

Hastie, T.J., Tibshirani, R.J., & Friedman, J.H. (2013). *The elements of statistical learning: Data mining, inference, and prediction.* New York, NY: Springer.

Hew, K.F., Cheung, W.S., & Ng, C.S.L. (2010, Nov.). Student contribution in asynchronous online discussion: A review of the research and empirical exploration. *Instructional Science*, *38*(6), 571–606.

Hewitt, J., Brett, C., & Peters, V. (2007). Scan rate: A new metric for the analysis of reading behaviors in asynchronous computer conferencing environments. *American Journal of Distance Education*, *21*(4), 215–231.

Joksimović, S., Gašević, D., Kovanović, V., Adesope, O., & Hatala, M. (2014, Jul.). *The Internet and Higher Education*, *22* (URL http://www.sciencedirect.com/science/article/pii/S1096751614000189).

Kanuka, H. (2011). Interaction and the online distance classroom: Do instructional methods effect the quality of interaction? *Journal of Computing in Higher Education*, *23*(2), 143–156.

Kuhn, D. (1995, May). Microgenetic study of change: What has it told us? *Psychological Science*, *6*(3), 133–139 (URL http://pss.sagepub.com.proxy.lib.sfu.ca/content/6/3/133).

Lust, G., Elen, J., & Clarebout, G. (2013a). Regulation of tool-use within a blended course: Student differences and performance effects. *Computers & Education*, *60*(1), 385–395.

Lust, G., Elen, J., & Clarebout, G. (2013b). Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. *Computers in Human Behavior*, *29*(5), 2013–2021.

Lust, G., Juarez Collazo, N.A., Elen, J., & Clarebout, G. (2012). Content management systems: Enriched learning opportunities for all? *Computers in Human Behavior*, *28*(3), 795–808.

Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., & Clarebout, G. (2011). Tool-use in a blended undergraduate course: In search of user profiles. *Computers & Education*, *57*(3), 2135–2144.

Macfadyen, L.P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599.

McGill, T.J., & Klobas, J.E. (2009). A task–technology fit view of learning management system impact. *Computers & Education*, *52*(2), 496–508.

Miles, J. (2001, Mar.). *Applying regression and correlation: A guide for students and researchers* (1st ed.). London; Thousand Oaks, Calif.: SAGE Publications Ltd.

Morris, L.V., Finnegan, C., & Wu, S. -S. (2005). Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education*, *8*(3), 221–231.

Murphy, E., & Coleman, E. (2004). Graduate students' experiences of challenges in online asynchronous discussions. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, *30*(2).

Nath, R., & Pavur, R. (1985). A new statistic in the one-way multivariate analysis of variance. *Computational Statistics & Data Analysis*, *2*(4), 297–315.

Penny, L., & Murphy, E. (2009). Rubrics for designing and evaluating online asynchronous discussions. *British Journal of Educational Technology*, *40*(5), 804–820 (URL http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8535.2008.00895.x/abstract).

Perkins, D.N. (1985). The fingertip effect: How information-processing technology shapes thinking. *Educational Researcher*, *14*(7), 11–17.

Phan, H.P. (2008). Unifying different theories of learning: Theoretical framework and empirical evidence. *Educational Psychology*, *28*(3), 325–340.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368–384.

Rourke, L., & Anderson, T. (2004). Validity in quantitative content analysis. *Educational Technology Research and Development*, *52*(1), 5–18.

Rourke, L., Anderson, T., Garrison, D.R., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. *The Journal of Distance Education*, *14*(2), 50–71.

Rovai, A.P. (2007). Facilitating online discussions effectively. *The Internet and Higher Education*, *10*(1), 77–88 (URL http://www.sciencedirect.com/science/article/pii/S1096751606000704).

Rubin, B., Fernandes, R., & Avgerinou, M.D. (2013). The effects of technology on the community of inquiry and satisfaction with online courses. *The Internet and Higher Education*, *17*, 48–57.

Sawyer, R.K. (2006). *The Cambridge handbook of the learning sciences.* Cambridge; New York: Cambridge University Press.

Schellens, T., Keer, H.V., Wever, B.D., & Valcke, M. (2007, Sep.). Scripting by assigning roles: Does it improve knowledge construction in asynchronous discussion groups? *International Journal of Computer-Supported Collaborative Learning*, *2*(2–3), 225–246 (URL: http://link.springer.com/article/10.1007/s11412-007-9016-2).

Schrire, S. (2006, Jan). Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. *Computers & Education*, *46*(1), 49–70 (URL: http://www.sciencedirect.com/science/article/pii/S0360131505000539).

Senko, C., Hulleman, C.S., & Harackiewicz, J.M. (2011). Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. *Educational Psychologist*, *46*(1), 26–47.

Shea, P., & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computers & Education*, *55*(4), 1721–1731.

Shen, D., Cho, M. -H., Tsai, C. -L., & Marra, R. (2013). Unpacking online learning experiences: Online learning self-efficacy and learning satisfaction. *The Internet and Higher Education*, *19*, 10–17.

Siemens, G., & Gasevic, D. (2012). Guest editorial — learning and knowledge analytics. *Educational Technology & Society*, *15*(3).

Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics.* Boston: Pearson/Allyn & Bacon.

Trigwell, K., & Prosser, M. (1991). Improving the quality of student learning: The influence of learning context and student approaches to learning on learning outcomes. *Higher Education*, *22*(3), 251–266.

Valle, R.D., & Duffy, T.M. (2009). Online learning: Learner characteristics and their approaches to managing learning. *Instructional Science*, *37*(2), 129–149.

Winne, P.H. (1982). Minimizing the black box problem to enhance the validity of theories about instructional effects. *Instructional Science*, *11*(1), 13–28.

Winne, P.H. (2006). How software technologies can improve research on learning and bolster school reform. *Educational Psychologist*, *41*(1), 5–17.

Winne, P.H., & Hadwin, A.F. (1998). Studying as self-regulated learning. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Metacognition in educational theory and practice. The educational psychology series*. (pp. 277–304). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Winne, P.H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, *27*(4), 551–572.

Wise, A.F., Speer, J., Marbouti, F., & Hsiao, Y. -T. (2013). Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviors. *Instructional Science*, *41*(2), 323–343.

Yen, J. -C., & Lee, C. -Y. (2011). Exploring problem solving patterns and their impact on learning achievement in a blended learning environment. *Computers & Education*, *56*(1), 138–145.

Zhou, M., & Winne, P.H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, *22*(6), 413–419.

4.3 Summary

As indicated by Akyol and Garrison (2011a), *"Critical thinking and inquiry is predicated upon an awareness and ability for learners to take responsibility and control to construct meaning and confirm knowledge."* (p. 183) As such, it is essential that our assessment and understanding of students' cognitive presence involves insights and information about students' self-regulation and metacognitive monitoring and control. Although the traces of student self-regulation are to some extent present in discussion transcripts, they are primarily concerned with the broader range of learning activities that students engage in, and as such are, for the most part, hidden from the external observers. This is acknowledged by Garrison et al. (2001) and the need for additional types of data to provide a richer assessment of cognitive presence, especially from the personal, reflective side of inquiry-based learning. The focus of this chapter is on the use of trace data of students use of available educational tools and resources to gain insights into their reflective dimension of inquiry-based learning and self-regulation. To the best of our knowledge, the work presented in this chapter is the first study which examined the cognitive presence from the "other" side of inquiry-based learning and not strictly focused on learning as expressed in the discourse transcripts.

With regards to the cognitive presence assessment model introduced in Chapter two (Figure 4 of the study by Kovanović, Gašević, Hatala, and Siemens 2017), the clustering system – like the classification system described in Chapter three – is represented on the top, assessment approach layer, with the primary source of data being the trace data of students' use of a learning platform. The clustering system utilizes the same student model as the classifier described in Chapter two. However, the clustering system is based on different task and evidence models (i.e., learning activities and quantitative measures of cognitive presence, respectively). Given the focus on measuring students' reflective learning activities through students' interactions with a broad range of tools and resources in a learning environment, the task and evidence models provide a more comprehensive list of learning activities and measures of student reflective learning, besides measures related to the use of online discussions.

As learning platforms vary considerably with respect to the trace data recordings, the clustering system presented in this chapter can be altered for the use in the different learning context than the one in the present study. For example, the Moodle learning management system adopted in the Kovanović, Gašević, Joksimović, et al. (2015) study does not provide information about time spent reading different discussion posts, only the information whether student accessed a discussion thread or not. However, certain learning platforms – such as Yellowdig [1] – provide an estimate of time spent reading each discussion message by examining how long the different parts of the discussion Web page were visible. This approach, which has already been used in the learning analytics research (Wise, Speer, Marbouti, & Hsiao, 2013; Wise, Hausknecht, & Zhao, 2014), could be used to provide important additional insights into students' course engagement and cognitive

[1] www.yellowdig.com

presence. To make use of these additional metrics of students' reading behavior, the only part of the clustering system that needs to be adjusted in the specification of the evidence model, most notably the list of extracted assessment metrics which are used to provide information about students' technology use. Similarly, if the course design is significantly different from the one described in the present study, the list of extracted quantitative metrics can be altered to reflect these changes. For example, if students use Wiki pages and personal blogs rather than traditional asynchronous discussion forums, then the list of extracted metrics would include measures which are related to the use of Wiki pages and personal blogs rather than online discussions. As such, the assessment approach presented here can be adjusted to accommodate for the particular unique characteristics of the learning environment or the course context.

5 Methodological challenges with trace data usage for learning analytics models

> Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.
>
> — John Tukey, *The Future of Data Analysis*

5.1 Introduction

As learning analytic involves processing of large amounts of learning data, an important part of learning analytics systems is the extraction of meaningful measures of student learning and engagement (Siemens et al., 2011). With this in mind, the educational research literature provides an important foundation for extraction of meaningful and pedagogically sound measures of student learning that can be then analyzed by different machine learning algorithms.

Although extracted quantitative measures depend on the type of analysis (e.g., sequence mining, social network analysis, classification, clustering, and topic modeling), the two most common type of quantitative measures being extracted are *count* of activities and *time-on-task* measures. The former represent the simplest measures that capture how many times a particular student executed a certain activity (e.g., the number of times the student posted a new message) while later capture the amount of time the student spent on a particular learning activity (e.g., the total time a student spent writing discussion messages). Time-on-task measures are directly based on the work of (Carroll, 1963) and (Bloom, 1974) who recognized the importance of active cognitive engagement and effort that student put into learning tasks on the eventual outcomes.

Given that modern Web-based online learning platforms only capture the discrete times when a particular list of events occurred (e.g., time when student opened a discussion, time when student posted a message), the time students' spent on various learning activities must be manually estimated from this, often incomplete, record of learning events. As such, there are several different approaches for extraction of time-on-task measures which can have a significant effect on the final results of an analytics-based system. Moreover, descriptions of learning analytics systems in the published literature often do not provide enough details regarding the ways in which particular measures were extracted, which brings important challenges for the generalizability of the reported research findings, as well as the applicability of the analytical systems in novel contexts.

The focus of this chapter is on the analysis of effects that different time-on-task estimation approaches have on the quality of the resulting learning analytics models, including the model of cognitive presence assessment. In particular, we compared how various time-on-task extraction procedures affect the result of statistical modeling of various student success measures. Of particular importance for this study is the examination how time-on-task estimation measures affect the modeling of cognitive presence and whether there are any particular estimation methods that provide better explanations for the observed levels of students' cognitive presence. As the effect of particular measures is also highly dependent on the adopted data mining algorithm, to keep the analysis simple and objective, we used simple regression models, as they are widely used and produce relatively stable results despite variations in the predictor variable estimates. Finally, we conducted an additional analysis to examine the effect of the time-on-task estimation methods when they are used together with count measures, which is very often done in practice.

The results of our study provide one of the first insights into the ways in which variability in input measures affect the final data analysis results. Our findings reveal that the adopted time-on-task estimation methods play a major role in shaping the overall analysis results, with as much as 23% of variance explained being solely accounted for an adopted estimation procedure. Interestingly, out of all assessed measures, cognitive presence was the most resilient to the way in which student time-on-task was estimated, with only 7% of variance explained being accounted for by the time-on-task estimation method. These findings have important implications for learning analytics, as they indicate that the sensitivity of models needs to be carefully considered to avoid models that do not generalize to the new context and previously unseen data.

5.2 Publication: Does time-on-task estimation matter? Implications on validity of learning analytics findings

The following section includes the verbatim copy of the following publication:

> Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., and Baker, R. S. (2015). Does time-on-task estimation matter? Implications on validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. doi:10.18608/jla.2015.23.6

Does Time-on-task Estimation Matter? Implications for the Validity of Learning Analytics Findings

Vitomir Kovanović
School of Informatics, University of Edinburgh, UK
v.kovanovic@ed.ac.uk

Dragan Gašević
Moray House School of Education and School of Informatics, University of Edinburgh, UK

Shane Dawson
Teaching Innovation Unit, University of South Australia, Australia

Srećko Joksimović
Moray House School of Education, University of Edinburgh, UK

Ryan S. Baker
Teachers College, Columbia University, USA

Marek Hatala
School of Interactive Arts and Technology, Simon Fraser University, Canada

ABSTRACT: With widespread adoption of Learning Management Systems (LMS) and other learning technology, large amounts of data — commonly known as trace data — are readily accessible to researchers. Trace data has been extensively used to calculate time that students spend on different learning activities — typically referred to as time-on-task. These measures are used to build predictive models of student learning in order to understand and improve learning processes. While time-on-task measures have been used in Learning Analytics research, the consequences of their use are not fully described or examined. This paper presents findings from two experiments regarding different time-on-task estimation methods and their influence on research findings. Based on modelling different student performance measures with popular statistical methods in two datasets (one online, one blended), our findings indicate that time-on-task estimation methods play an important role in shaping the final study results, particularly in online settings where the amount of interaction with LMS is typically higher. The primary goal of this paper is to raise awareness and initiate debate on the important issue of time-on-task estimation within the broader learning analytics community. Finally, the paper provides an overview of commonly adopted time-on-task estimation methods in educational and related research fields.

Keywords: Time-on-task, measurement, learning analytics, higher education, Learning Management System (LMS), Moodle

1 INTRODUCTION

A main precondition for the adoption of learning analytics is the collection of relevant data about student learning. One widely used type of data is *trace data* about student interactions within a Learning Management System (LMS). These trace data typically take the form of *event streams,* timed lists of events performed through system use, typically by either students (e.g., reading discussions, submitting assignments) or instructors (e.g., uploading student grades). One benefit of trace data is that it can be easily converted to aggregate numerical *count data* showing frequencies of different actions for each student. Count data is useful in the educational context as it enables an overview of student learning activities and provides the opportunity to develop a broad range of predictive models of student performance and student monitoring systems.

In addition to the use of count data, LMS trace data has been extensively used to estimate students' actual time spent online as a proxy of academic activity and learning. Beginning with early studies of traditional classroom learning in the 1970s, the amount of time students actually spent on learning has been identified as one of the central constructs affecting learning success (Bloom, 1974; Stallings, 1980). To this day, one of the primary ways of improving student learning is to develop learning activities that support longer engagement periods with course content or peers (Stallings, 1980). Instead of using count measures, time-on-task measures provide a more "accurate" estimate of the amount of effort students spend learning.

Despite time-on-task being identified as an important measure of student learning, its accurate estimation is a non-trivial task (Karweit & Slavin, 1982). Given the typical client-server architecture of Web applications and the fact that most learning systems only record streams of important system events, a reconstruction of times spent on different learning activities is required. Typically, the estimation process involves measuring time differences between subsequent events in the event stream as the more fine-grained information is often not available. The challenge with this approach is that between two event-stream activity records students often engage in some other activities not related to their learning. For example, a student may be studying in the evening and then continue their learning session the following morning. In that case, the time span between the last learning activity in the evening and the first learning activity in the morning would be very long, and therefore affect the accuracy of naïve time-on-task estimation methods that do not take into the account these situations.

While it is an important part of data collection, the estimation of time-on-task measures is rarely discussed in detail within learning analytics research. Typically, researchers adopt a heuristic approach (e.g., limit all activities to 10, 30, or 60 minutes) (Ba-Omar, Petrounias, & Anwar, 2007; Munk & Drlík, 2011) and do not address the consequences of such adopted heuristics on the produced statistical model. In this paper, we try to evaluate what are the consequences of the different estimation heuristics on the results of the final predictive model. More precisely, we looked at how different strategies for time-on-task estimation affect the results of several multiple linear regression models in two separate datasets from fully online and blended courses. In order to provide a more comprehensive analysis as an outcome measure in the predictive models, we used students' final grades, individual assignment grades, discussion participation grades, and number of messages with higher levels of cognitive presence — a central component of a widely used Community of Inquiry model (CoI) of distance education (Garrison, Anderson, & Archer, 1999, 2001). Based on the findings of the present study, we offer some practical guidelines for improving the validity of research in learning analytics. We also suggest greater attention to this topic in future learning analytics research.

2 BACKGROUND

2.1 Time-on-task in Educational Research

2.1.1 Origins of time-on-task in educational research

There is a long tradition for the use of time in education research (Bloom, 1974). In 1963, Carroll proposed a model of learning where time was a central element, and learning was defined as a function of the effort spent in relation to the effort needed. Carroll, however, made a distinction between *elapsed time* and the time students *actually spend on learning* (1963). Student learning depends on how the time is used, not the total amount of time allocated (Stallings, 1980). There has been extensive research in the 1970s noting the benefits of increased learning time on overall learning quality (Karweit, 1984; Karweit & Slavin, 1982; Stallings, 1980). In this context, an increase in time-on-task was considered one of the key principles of effective education (Chickering & Gamson, 1989).

A main challenge with research on the effects of time on learning is different operationalizations of the time-on-task construct (Karweit & Slavin, 1982). Some researchers (e.g., Helmke, Schneider, & Weinert, 1986; Cohen, Manion, & Morrison, 2007) used typical observational methods such as monitoring student behaviour at specified time intervals and coding that behaviour using a predefined coding scheme. Others (e.g., Admiraal, Wubbels, & Pilot, 1999) adopted very different and cruder notions of time-on-task, such as number of lectures attended, number of school days in a year, or hours in a school day. As pointed out by Karweit and Slavin (1982), differences in definitions of on-task and off-task behaviour, observation intervals, and sample sizes led to important inconsistencies in this research domain. According to Karweit (1984), the interpretation of significant findings related to time-on-task measures requires careful examination and caution.

2.1.2 Recent studies of student time-on-task

Despite prior warnings by Karweit and Slavin (1982) regarding time-on-task estimation, recent empirical studies (Calderwood, Ackerman, & Conklin, 2014; Judd, 2014; Rosen, Mark Carrier, & Cheever, 2013) continue to illustrate the complexities and possible inaccuracies linked to time estimation in the digital age. Given the ubiquitous access to technology, student learning activities are characterized by high levels of distraction and multi-tasking, which are shown to have negative effects on student attention and learning (Bowman, Waite, & Levine, 2015). For example, Calderwood et al. (2014) conducted a laboratory study with 58 participants that looked at their levels of distraction over a three-hour period of self-directed learning using various observational techniques (i.e., eye-tracking, surveillance camera, and video recorder). The striking finding is that even in the “sterile” and controlled laboratory environment students engaged, on average, in 35 distractions (of six seconds or more) with a total distraction time of 25 minutes (Calderwood et al., 2014). Similar results were found by Judd (2014), who looked at the levels of student multi-tasking while engaged in a learning activity. Using a specifically designed tracing application installed on the computers of 1,249 participants, Judd noted that Facebook users spent almost 10% of their study time on Facebook rather than studying. In addition, 99% of student study sessions involved some form of multi-tasking. Finally, the Rosen et al. (2013) field observational study of 263 participants looked at students' learning behaviour over a 15-minute study period and found, on average, that students spent only 10 of 15 minutes engaged in learning and were capable of maintaining only six minutes of on-task behaviour.

The above research sheds some light on the study habits of learners in the digital age. Whatever "correct" distraction times may be, it is certain that today's students are engaging in much more multi-tasking and off-task behaviours that affect the accuracy of measuring student time-on-task. We should note that in this context "off-task" should be understood as "off-system" meaning that students spend some time outside the system. This does not necessarily mean not engaging in productive learning activities (e.g., reading a printed document or attending a study group meeting); however, given that time-on-task estimates are used to understand learning activities and often to build predictive models of student success or identify students at risk, there is a need to provide better estimates of students' time-on-task. In this context, there is a further imperative for researchers to account for these off-system activities and off-task distractions when determining time-on-task estimations through trace data. It is very likely that similar levels of distraction are present in many of the datasets that learning analytics researchers use in their studies. With this in mind, the goal of the present study is to examine what effects different techniques for calculating time-on-task from LMS trace data have on the results of final learning analytics models.

2.1.3 Time-on-task and learning technology

The previously described observational techniques have also been used in many studies (Baker, Corbett, Koedinger, & Wagner, 2004; Smeets & Mooij, 2000; Worthen, Van Dusen, & Sailor, 1994) for examination of student behaviour and time-on-task analysis when working with educational technology. For example, research in the domain of Intelligent Tutoring Systems (ITS) has sought to identify off-task behaviour and its effects on learning (Baker et al., 2004; Baker, 2007; Cetintas, Si, Xin, & Hord, 2010; Cetintas, Si, Xin, Hord, & Zhang, 2009; Pardos, Baker, San Pedro, Gowda, & Gowda, 2013; Roberge, Rojas, & Baker, 2012).

The adoption of educational technology has enabled relatively easy calculation of student time-on-task based on the trace data collected by the software system. While this approach has been adopted in many research studies (Grabe & Sigler, 2002; Kraus, Reed, & Fitzgerald, 2001), the details of the process are not always described. While some of these studies (Grabe and Sigler, 2002) described the challenges that the process of time-on-task estimation entails, most of the studies do not. In their study, Grabe and Sigler (2002) used several heuristics for time-on-task estimation: 1) all learning actions longer than 180 seconds were estimated to be 120 seconds long, 2) all multiple choice answering actions to be at maximum 90 seconds, and 3) last actions within each study session were estimated at 60 seconds.

More recent research in the ITS field has led to the development of several machine learning systems for automated detection of student off-task behaviour based on trace data (Baker, 2007; Cetintas et al., 2010; Cetintas et al., 2009). The development of such models was made possible due to the availability of field observational data, thereby providing a "gold standard" for testing the performance of different models. In his study, Baker (2007) identified a time of 80 seconds to be the best cut-off threshold for identification of off-task behaviour. The best performing model for off-task behaviour detection also made use of a broader range of features, with a particularly useful feature being the standardized difference in duration among subsequent actions (i.e., very fast action followed by a very slow action or vice versa). This research provides an empirical analysis of the different approaches for detection of off-task behaviour and lays the groundwork for reproducible and replicable research in the ITS field.

2.2 Web-Usage Mining

2.2.1 Process & heuristics

User activities are extensively analyzed in the area of Web Usage Mining (WUM) (Cooley, Mobasher, & Srivastava, 1997), which is "the automatic discovery of user access patterns from Web servers" (Cooley et al., 1997, p. 560). Data pre-processing is recognized as a crucial step in WUM analysis (Cooley et al., 1997; Hussain, Asghar, & Masood, 2010; Munk & Drlík, 2011; Munk, Kapusta, & Švec, 2010) and is estimated to take typically between 60% and 80% of the total analysis time (Hussain et al., 2010; Marquardt, Becker, & Ruiz, 2004).

Typically, web-usage mining involves the analysis of *clickstream data* being recorded as users navigate through different parts of a Web-based system. According to Chitraa and Davamani (2010), the pre-processing in WUM consists of four separate phases: 1) *Data cleaning*, which involves removal of irrelevant log records; 2) *User identification*, typically based on their IP addresses and Web user agent resolution; 3) *Session identification*, with the goal of splitting user access information into separate system visits; and 4) *Path completion*, which deals with issues of missing information in the server access log (e.g., due to caching by proxy servers). Of direct importance for the studies presented in this paper is the notion of different strategies for session identification:

1. *Time-oriented heuristics*, which place an upper limit on the total session time (typically 30 minutes), or an upper limit on a single Web page time (typically 10 minutes) (Cooley, Mobasher, & Srivastava, 1999; Mobasher, Cooley, & Srivastava, 1999). Early empirical studies found 25.5 minutes to be an average duration of Web session (Catledge & Pitkow, 1995).
2. *Navigation-oriented heuristics*, which look at web page connectivity to identify user sessions. When for the same IP address two consequent pages in the access log are not directly linked, then this signals the start of a new user session.

As indicated by Chitraa and Davamani (2010), time-oriented heuristics are simple, but often unreliable, as users may undertake parallel off-task activities. Hence, it can be problematic to define user sessions based on time. Munk et al. (2010) adopted 10-minute timeout intervals for session identification and identified path completion pre-processing as an important step for improving the quality of extracted data. Similarly, Raju and Satyanarayana (2008) proposed a complete pre-processing methodology and suggested the use of 30-minute session timeout intervals.

2.2.2 Web usage mining in distance education

With the transition to Web-based learning technologies and with the broader adoption of LMS systems, several researchers (e.g., Ba-Omar et al., 2007; Marquardt et al., 2004) have adopted traditional WUM techniques to analyze learning data. It is important to note that certain characteristics of LMS systems make the process somewhat simpler. For example, user identification is trivial, as all learning platforms require a student login (Marquardt et al., 2004; Munk & Drlík, 2011). Likewise, modern LMS systems (e.g., Moodle) store student activity information in their relational databases, and therefore typical WUM analysis of LMS data does not require the analysis of plain Web server logs, which simplifies the data cleaning process (Munk & Drlík, 2011).

In the learning contexts, one of the earliest studies that addressed student time-on-task is by Marquardt, Becker, and Ruiz (2004). Their approach is unique in offering a different conceptualization of user session. Essentially, the authors use *reference session* to indicate a typical user session, and *learning session* to indicate a user session spanning multiple days and focusing on a particular learning activity. For identification of reference sessions Marquardt et al. (2004) also recommend using timeout intervals, but they do not provide a recommendation on a particular timeout value. This approach is used in many WUM studies of learning technologies, such as Ba-Omar et al. (2007) and Munk and Drlík (2011) who used 30- and 15-minute session timeouts, respectively.

In addition to the work drawing on research from Web mining, there are also more recent studies from the fields of learning analytics (LA) and educational data mining (EDM) that adopt novel strategies to address the issues of time-on-task estimation. For example, the study by del Valle and Duffy (2009) reported the use of a 30-minute timeout interval to detect the end of user sessions, and for each session estimated the duration of last action as an average time spent on a given action by a particular user. Del Valle and Duffy (2009) point out that the estimation of student time-on-task based on trace data is made under the assumption that time between two logged events is spent on learning — and that similar assumptions are made in the research of other learning modalities.

In a similar manner Wise, Speer, Marbouti, and Hsiao (2013) examined the distribution of action durations and used a 60-minute inactivity period as an indicator of the end of user activity. The last action of each session is estimated based on the length of the particular message and the average speed at which the user was conducting a particular action (i.e., reading, posting, or editing a message). In the context of mining trace data from collaborative learning environments, Perera, Kay, Koprinska, Yacef, and Zaiane (2009) used a time-based heuristic to define activity sessions using a 7-hour inactivity period.

There are also many studies in the LA and EDM fields that do not discuss and report details of how time-on-task measures were calculated (e.g., Lust, Elen, & Clarebout, 2013a, 2013b; Lust, Vandewaetere, Ceulemans, Elen, & Clarebout, 2011; Macfadyen & Dawson, 2010; Romero, Espejo, Zafra, Romero, & Ventura, 2013; Romero, Ventura, & García, 2008; Wise, Zhao, & Hausknecht, 2013). Typically, those studies make use of both count and time-on-task measures. As such, it would appear likely that researchers used time differences from the raw data or simple time-based heuristics such as the ones described above.

Several researchers have adopted unique techniques for time-on-task estimation. For example, Brown and Green (2009) calculated time spent reading discussions by extracting the average number of words per discussion and then multiplying it by 180 words per minute (which was obtained empirically). The challenge with this approach is in its inability to detect shallow reading and skimming (i.e., reading that is faster than 6.5 words per second) (Hewitt, Brett, & Peters, 2007), as done in similar studies (Oztok, Zingaro, Brett, & Hewitt, 2013; Wise, Speer, et al., 2013; Wise, Zhao, et al. 2013b) that estimated time-on-task from trace-data. Some studies also used self-reported data on the amount of time students spent using the system (e.g., García-Martín & García-Sánchez, 2013; Hsu & Ching, 2013; Romero & Barbera, 2011), and this approach raises an additional set of reliability challenges (Winne & Jamieson-Noel, 2002). Finally, in laboratory settings, Guo, Wang, Moore, Liu, and Chen (2009) and Kolloffel, Eysink, and Jong (2011) measured time-on-task as the difference between the start and the end of an experimental learning activity.

3 RESEARCH QUESTIONS: EFFECTS OF TIME-ON-TASK MEASURING ON ANALYTICS RESULTS

Although time-on-task measures from LMS trace data have been used extensively in learning analytics research, to the best of our knowledge there have been no studies that address the challenges and issues associated with their estimation and that investigate what effects the adopted estimation methods have on the resulting analytical models. The primary goal of this paper is to raise awareness in the learning analytics research community about the important implications that adopted estimation methods have. Thus, the main research question for this study is this:

> What effects do different methods for estimation of time on-task-measures from LMS data have on the results of analytical models? Are there differences in their statistical significance and overall conclusions that can be drawn from them?

In order to provide a comprehensive overview of the effect that time-on-task estimation has on study results, it is equally important to acknowledge the specifics of each individual course. Given that students' behaviour, conceptions of learning, and the use of learning systems are all highly dependent on the particular course specifics (e.g., course design, organization, subject domain) (Cho & Kim, 2013; Gašević, Dawson, Rogers, & Gašević, 2015; Trigwell, Prosser, & Waterhouse, 1999), the second goal of our study is to investigate how differences between the courses moderate the effects of different time-on-task estimation methods. Hence, our second research question is this:

> Are the effects of time-on-task estimation consistent across the courses from different subject domains and with different course organizations? Is there an association between the level of LMS use and the effect of time-on-task estimation strategies?

The majority of studies incorporating time-on-task estimation provide insufficient details concerning the adopted procedures and measurement heuristics, which are necessary to replicate their research findings. As the adopted techniques may have significant effects on the results of published studies, the learning analytics community should be cautious about interpreting any results that involve time-on-task measures from LMS data.

4 STUDY DATASETS

4.1 Online Course Dataset

4.1.1 Course organization

The first dataset is from a 13-week-long masters-level fully online course in software engineering offered at a Canadian public university. Given its postgraduate level, the course was research intensive and focused on contemporary trends and challenges in the area of software engineering. The course used the university's Moodle platform (Moodle HQ, 2014), which hosted all resources, assignments, and online discussions for the course. This particular course was selected because it was a fully online course with strong emphasis on the use of the LMS platform in particular assignments, resources, and forum Moodle components — also known as Moodle system modules. To finish the course successfully students were expected to complete several activities including four tutor-marked assignments (TMAs):

- **TMA1** (15% of the final grade): Students were requested to 1) select and read one peer-reviewed paper, 2) prepare a video presentation for other students describing and analyzing the selected paper, and 3) make a new discussion thread in the online forums where students would discuss each other's presentations.
- **TMA2** (25% of the final grade): Students were required to write a literature review paper (5–6 pages in the ACM proceedings format) on a particular software engineering topic. The mark for this assignment was determined as follows: 1) 80% based on two double-blind peer reviews (each contributing 35% of the paper grade) and the instructor review (contributing 30% of the paper grade), and 2) 20% given by the instructor based on the quality of the peer-review comments.
- **TMA3** (15% of the final grade): Students were requested to demonstrate critical thinking and synthesis skills by answering six questions (400–500 words each) related to the course readings.
- **TMA4** (30% of the final grade): Students were required to work in groups of 2–3 on a software engineering research project. The outcome was a project report along with a set of software artefacts (e.g., models and source code) marked by the instructor.

- **Course Participation** (15% of the final grade): Students were expected to participate productively in online discussions for the duration of the course.

The data was obtained from Moodle's PostgreSQL database and consisted of 167,000 log records produced by 81 students who completed the course, which was offered six times: Winter 2008 (N=15), Fall 2008 (N=22), Summer 2009 (N=10), Fall 2009 (N=7), Winter 2010 (N=14), and Winter 2011 (N=13). During the course, students produced 1,747 discussion messages that were also used as an additional dataset for this study. Table 1 shows the detailed description of each course offering used in this study.

4.1.2 Extraction of count and time-on-task measures

From the collected trace data, we extracted five count measures, shown in Table 2, and corresponding time-on-task measures using different estimation strategies, which will be covered in detail in the Methodology section. The extracted measures correspond to the activities in which the students were expected to engage. The count measures were easily extracted from Moodle trace data, as the number of times each action is recorded for every student. Similarly, time-on-task measures were extracted as the total amount of time each student spent on a particular type of activity.

4.1.3 Extraction of performance measures

In addition to count measures, we extracted a set of four academic performance measures: 1) TMA2 grade, 2) TMA3 grade, 3) course participation grade, and 4) final course percent grade. We decided to use TMA2, TMA3, and course participation grades since they stipulated a high use of the LMS system, while the other two assignments (TMA1 and TMA4) expected more "offline" work from the students. Finally, given that many studies examined the relationship between final course grades and student use of LMSs, we included final course grade as an additional "high-level" measure of academic performance.

Table 1: Online course dataset: Course offering statistics

	Students	Actions	Messages	Actions/Student	Messages/Student
Winter 2008	15	33,976	212	2,265	14.1
Fall 2008	22	49,928	633	2,269	28.8
Summer 2009	10	21,059	243	2,106	24.3
Fall 2009	7	11,346	63	1,621	9.0
Winter 2010	14	31,169	359	2,226	25.6
Winter 2011	13	19,783	237	1,522	18.2
Average (SD)	13.5 (5.1)	27,877 (13,561)	291.2 (192.4)	2,002 (340)	20.0 (7.6)
Total	81	167,261	1,747		

Table 2: Online course dataset: Extracted measures

	Count Measures		
#	Module	Name	Description
1	Assignment	AsignmentViewCount	Number of assignment views.
2	Forum	ResourceViewCount	Number of resources views.
3	Forum	DiscussionViewCount	Number of course discussion views.
4	Forum	AddPostCount	Number of posted messages.
5	Forum	UpdatePostCount	Number of post updates.
	Time-on-Task Measures		
#	Module	Name	Description
1	Assignment	AsignmentViewTime	Time spent on course assignments.
2	Forum	ResourceViewTime	Time spent reading course resources.
3	Forum	DiscussionViewTime	Time spent viewing course discussions.
4	Forum	AddPostTime	Time spent posting discussion messages.
5	Forum	UpdatePostTime	Time spent updating discussion messages.
	Performance Measures		
#	Name		Description
1	TMA2Grade		Grade for literature review paper.
2	TMA3Grade		Grade for journal papers readings.
3	ParticipationGrade		Grade for participation in course discussions.
4	FinalGrade		Final grade in the course.
5	CoIHigh		Integration and resolution message count.

In order to provide a more comprehensive experimental setting that includes several types of dependent measures, we used an additional set of measures based on the popular Community of Inquiry (CoI) framework (Garrison et al., 1999). We selected the CoI model because it was the basis for the design of the target course (cf. Gašević, Adesope, Joksimović, & Kovanović, 2015). Furthermore, the CoI framework is one of the most well researched and validated models of distance education (cf. Swan & Ice, 2010) that defines important dimensions of online learning and offers a coding instrument for measurement (Garrison et al., 1999) of these dimensions. In the present study, we focused on the *cognitive presence* construct, which describes students' development of critical and deep thinking skills as consisting of four phases: 1) *Triggering event*, 2) *Exploration*, 3) *Integration*, and 4) *Resolution*. Early research (Garrison et al., 2001) has indicated that a majority of students do not easily nor readily progress to the later stages of cognitive presence. With the intention of examining association between different time-on-task measures and development of cognitive presence, we extracted one additional performance measure, `CoIHigh`, namely, the number of messages in integration and resolution phases. We coded discussion messages using the CoI coding scheme for cognitive presence described by Garrison et al. (2001). Each message was coded by two human coders who achieved an excellent inter-rater agreement (Cohen's kappa=.97), disagreeing on only 32 messages. The results of the coding process are shown in Table 3.

Table 3: Message coding results

ID	Phase	Messages	(%)
0	Other	140	8.01%
1	Triggering Event	308	17.63%
2	Exploration	684	39.17%
3	Integration	508	29.08%
4	Resolution	107	6.12%
	All Phases	1,747	100%

4.2 Blended Courses Dataset

4.2.1 Courses organization

In order to examine the effects of diverse course organizations on the use of different time-on-task estimation strategies, we used a large dataset from a Spring 2012 offering of nine first-year courses at a large Australian public university. All nine courses were part of the university-wide student retention project called Enhancing Student Academic Potential (ESAP). The project was organized and coordinated by the university's central learning and teaching unit to provide support for first-year students identified as having learning behaviours that tended to lead to suboptimal academic success. Participation in ESAP was based on a consistent low retention in the program and course success in the past five years. In addition, all ESAP courses were required to have more than 150 students enrolled. Before the start of the courses, all students were informed about compliance with the university's ethics and privacy regulations and that the LMS data would be collected and used for improving the quality of the courses and understanding of student learning behaviours.

All nine courses were offered using a blended learning approach in which face-to-face instruction was accompanied by an online component provided by the university's central Moodle LMS platform (e.g., assignments, resources, quizzes, chat, student discussions). The nine courses of the ESAP initiative included in this study were from a wide range of disciplines. Those include two courses from biology (BIOL 1 and BIOL 2), and one course from accounting (ACCT), communications (COMM), computer science (COMP), economics (ECON), graphics design (GRAP), marketing (MARK), and mathematics (MATH). The general information about the size of each course's data is shown in Table 4. In total, the dataset consisted of slightly more than 4,000 students that generated 4.6 million action records and about 3,000 discussion messages. On average, each course had 449 students (SD=243) and a little over 250,000 relevant LMS trace records.

4.2.2 Extraction of count, time-on-task, and performance measures

As with a fully online dataset, the data for each course included only students that completed the course and included only the ones that were relevant from the standpoint of course organization. As each course had different organization and different expectations for LMS use, we included only the data aligned with course organization. The usage summary for different Moodle modules (e.g., discussions, assignments, quizzes, chat) in each course is shown on Table 5. As we can see, most courses adopted assignments, forums, resources, and turnitin modules, while a smaller number of courses used other modules.

We extracted trace data for activities that students were expected to use by course design and were related to learning, similarly to the first dataset. As most Moodle modules have actions not corresponding to learning activities (e.g., listing all discussions or listing all assignments), from each of the modules we focused only on actions related to student learning. Finally, for certain actions — such as forum search — there is no meaningful notion of time, so in those cases we extracted only count measures. The complete list of extracted measures is shown in Table 6. We extracted six measures that do not have a corresponding time measure, and 13 measures that had meaningful corresponding time-on-task measures. As measures related to the number of discussion message edits (i.e., `UpdatePostCount` and `UpdatePostTime`) were close to zero in all nine courses, we removed those measures from our further analysis. A detailed overview of extracted count measures for each course is given in Table 7. As we can see, courses differed in their volume of activity, and mostly made use of all activities defined by the course design. The only notable exceptions were COMP and GRAP courses that did not make use of online discussions, even though they were made available — but not directly scaffolded — by the course design.

In contrast to the first dataset, in which we extracted a variety of outcome measures, for the second analysis we focused only on a single outcome measure, a course final percentage grade. Given that each course has a specific grading structure and list of assignments, in order to examine the effect of course organization we focused on the outcome measure common to all courses — course final grade. This enabled us to see the differences in results of regression analyses between courses across different time-on-task estimation approaches.

Table 4: Blended courses dataset: Course statistics

Course	Students	Actions	Messages	Actions/Students	Messages/Students
ACCT	734	327,423	515	446	0.70
BIOL 1	216	221,102	206	1,024	0.95
BIOL 2	648	595,730	1024	919	1.58
COMM	494	210,085	509	425	1.03
COMP	236	100,638	0	426	0.00
ECON	646	409,116	416	633	0.64
GRAP	172	14,746	0	86	0.00
MARK	712	327,144	407	459	0.57
MATH	191	119,997	56	628	0.29
Average (SD)	449 (243)	258,442 (172,570)	348 (329)	561 (282)	0.64 (0.51)
Total	4,049	4,651,962	3,133		

Table 5: Blended courses dataset: Course module usages

	ACCT	BIOL 1	BIOL 2	COMM	COMP	ECON	GRAP	MARK	MATH
Assignment	X	X		X	X	X		X	X
Book	X		X			X			
Chat								X	
Course Logins	X	X	X	X	X	X	X	X	X
Feedback			X						
Forum	X	X	X	X	X	X	X	X	X
Gallery	X								
Map			X						
Quiz		X	X		X	X			
Resource	X	X	X	X	X	X	X	X	X
Turnitin	X			X	X	X		X	X
Virtual Classroom			X						

Table 6: Blended courses dataset: Extracted measures

#	Module	Name	Description
	Count-only Measures (no corresponding time-on-task measure)		
#	Module	Name	Description
1	Assignments	AssignmentUploadCount	Number of assignment uploads.
2	Book	BookPrintCount	Number of book printings.
3	Course	CourseViewCount	Number of course homepage views.
4	Feedback	FeedbackCount	Number of feedbacks submitted.
5	Forum	ForumSearchCount	Number of forum searches.
6	Turnitin	TurnitinSubmissionCount	Number of turnitin submissions.
	Count Measures (with corresponding time-on-task measure)		
#	Module	Name	Description
1	Assignments	AssignmentViewCount	Number of assignment views.
2	Book	BookViewCount	Number of book views.
3	Chat	ChatViewCount	Number of chat views.
4	Chat	ChatTalkCount	Number of chat messages.
5	Forum	ViewDiscussionCount	Number of forum discussion views.
6	Forum	AddPostCount	Number of forum messages written.
7	Gallery	GalleryViewCount	Number of gallery views.
8	Map	MapViewCount	Number of geo map views.
9	Quiz	QuizViewCount	Number of quiz views.
10	Quiz	QuizAttemptCount	Number of quiz attempts.
11	Quiz	QuizReviewCount	Number of quiz reviews.
12	Resources	ResourceViewCount	Number of course resource views.
13	Virtual classroom	AdobeConnectViewCount	Number of virtual classroom views.
	Time-on-Task Measures (with corresponding count measures)		
#	Module	Name	Description
1	Assignments	AssignmentViewTime	Time spent viewing assignments
2	Book	BookViewTime	Time spent viewing course books.
3	Chat	ChatViewTime	Time spent viewing chat records.
4	Chat	ChatTalkTime	Time spent entering chat messages.
5	Forum	ViewDiscussionTime	Time spent viewing discussions.
6	Forum	AddPostTime	Time spent writing forum messages.
7	Gallery	GalleryViewTime	Time spent viewing course galleries.
8	Map	MapViewTime	Time spent viewing geo maps.
9	Quiz	QuizViewTime	Time spent viewing course quizzes.
10	Quiz	QuizAttemptTime	Time spent doing course quizzes.
11	Quiz	QuizReviewTime	Time spent reviewing quiz results.
12	Resources	ResourceViewTime	Time spent viewing resources.
13	Virtual classroom	AdobeConnectViewTime	Time spent in virtual classroom.
	Performance Measures		
#	Name		Description
1	FinalGrade		Final percent grade in the course.

Table 7: Blended courses dataset: Course actions counts

	ACCT	BIOL 1	BIOL 2	COMM	COMP	ECON	GRAP	MARK	MATH
StudentCount	734	216	648	494	236	646	172	712	191
Avg.Grade	72.7 (140.2)	60.4 (68.1)	74.5 (123.4)	85.5 (163.3)	82 (137)	73.2 (134.1)	64.7 (73.1)	74.4 (122.2)	69.2 (119.7)
Assign.UploadCount	2 (2.8)	0 (0)		7.4 (5.1)	2.8 (2.5)	7 (4.5)		5.1 (2.4)	2.4 (3.7)
Assign.ViewCount	6.7 (8.6)	21.4 (11.5)		27.3 (19)	11.7 (8.1)	30.1 (18.2)		23.5 (14.4)	23.9 (13.5)
BookViewCount	4.8 (6.8)		5.2 (8)			2.1 (2.1)			
BookPrintCount	0 (0.1)		0.1 (0.8)			0.1 (0.3)			
ChatTalkCount								0.2 (2.6)	
ChatViewCount								0.4 (1.1)	
CourseViewCount	58.5 (63)	125 (76.2)	135.4 (114.9)	60.8 (49)	71.7 (49.2)	84.5 (70.6)	11.2 (9.2)	59 (46.2)	98.7 (62.4)
FeedbackCount			0.7 (0.8)						
ForumSearchCount	0.7 (4.9)	0.1 (0.6)	0.1 (0.4)	0.1 (0.9)	0 (0)	0.1 (1.3)	0 (0)	0.1 (0.6)	0.1 (0.6)
ViewDisc.Count	27.9 (62.6)	37.4 (36.3)	36.8 (77.3)	43.4 (61.5)	0 (0)	30 (42)	0 (0)	22 (33.9)	11.5 (14.1)
AddPostCount	0.3 (2.5)	0.6 (2.5)	1.1 (4.2)	0.6 (2)	0 (0)	0.4 (1.3)	0 (0)	0.3 (1.6)	0.1 (0.6)
GalleryViewCount	0.9 (1.6)								
MapViewCount			0.4 (1.2)						
QuizViewCount		29.7 (15.6)	51.3 (59.9)		3.1 (6.7)	6.8 (12.3)			
QuizAttemptCount		8.1 (2.2)	30.7 (36.6)		0.7 (1.6)	3.2 (6)			
QuizReviewCount		19.4 (51.5)	30.5 (37.2)		1.8 (5.7)	3.5 (8.9)			
Res.ViewCount	45.9 (62.7)	71.6 (41.9)	137.8 (91)	23.2 (14.6)	0.6 (0.8)	60.2 (101.9)	11.1 (10.5)	54.8 (40.3)	92.3 (63.6)
TurnitinSub.Count	0.9 (1)			3.4 (1.9)	2.2 (1.4)	3.2 (1.7)		2.5 (1.1)	1 (1.6)
AdobeCon.ViewCount			12.4 (24.7)						

5 METHODOLOGY

5.1 Extraction of Time-on-task Measures

5.1.1 Time-on-task extraction procedure

In order to calculate time-on-task measures we processed trace data available in the Moodle platform. Table 8 shows a typical section of the logged data. Moodle itself does not record the duration of each individual action, but rather stores only timestamps of important "events" completed by the students or the system. Thus, in order to calculate the time spent on different activities, a difference between subsequent log records is measured. For example, to calculate time spent viewing discussion D1, we calculated the difference between its start time and the start time of the following activity in the log (T2–T1). This is the simplest, most straightforward way of determining time-on-task calculations.

As some of the logged actions have unique properties, they require special attention. For example, a certain number of logged activities are instantaneous and cannot be attributed to a meaningful duration of time (e.g., marking discussion as read, or performing a search in discussion boards). Thus, the time periods between these actions and subsequent actions should be added to time-on-task estimates of *preceding* actions in the action log. For example, in Table 8, time spent viewing discussion D2 should — besides period T2–T3 — also include period T3–T4 as the user continued to read the same discussion after marking it as read. Thus, the total time-on-task for viewing discussion D2 should be calculated as T4–T2.

Table 8: Typical trace data. Blue cursive indicates actions with overestimated time-on-task, while red boldface indicates actions that require special non-standard calculation of time-on-task

Time	User	Action	Duration
…	…	…	…
T0	User U	UserLogin	0s
T1	User U	Start Viewing Discussion D1	T2 – T1
T2	User U	Start Viewing Discussion D2	**T4** – T2
T3	User U	Mark Discussion D2 as Read	T4 – T3
T4	User U	Start Viewing Discussion D3	0s
T5	User U	Submit New Message M1	**T5 – T4**
T6	User U	Start Viewing Discussion D4	*T7* – t6
…	…	prolonged time period	…
T7	User U	Start Viewing Assignment TMA1	T8 – T7
T8	User U	Start Viewing Resource R1	*T9* – T8
…	…	prolonged time period	…
T9	User U	User Login	T10 – T9
T10	User U	Start Viewing Resource R2	T11 – T10
T11	User U	Start Viewing Discussion D5	*T12* – T11
T12	User U	User Login	T13 – T12
…	…	…	…

It is also important to note that Moodle records certain actions at their end, rather than their start. In these instances, a "backward" time-on-task estimation is required. This is best illustrated through an example from Table 8 where student `U` starts viewing discussion `D3` at time `T4`. After a while, the student clicks the "Post Reply" button to post his response to the discussion. A pop-up dialog for writing a new message appears and the student starts typing his response. However, *Moodle does not record the start of the message writing*. It is only *after* the student presses the "Submit" button, that an action is logged by the system (time T5). Thus, the time spent writing the message should be calculated "backwards," as T5–T4. Given that the exact moment when the student started writing his response is not recorded, it is also not possible to tell how much time the student actually spent writing the response and how much on reading the discussion prior to writing the response. Thus, time spent reading discussions preceding a reply by a student could not be precisely determined from the current format of Moodle logs. This is a particular challenge of the Moodle platform that should be considered when calculating time-on-task estimates from Moodle trace data.

5.1.2 Two challenges of time-on-task estimation

An important characteristic of Moodle relates to the way in which user sessions are handled. Typically, a student session is preserved as long as the student's browser window is open. Thus, if the student stops using the system and engages in an alternate activity, it would be impossible to detect the off-task behaviour based on Moodle logs alone. A typical solution for dealing with such cases is to use some form of time-based heuristic — as described in Section 2 — and place a maximum value on the duration of activities (usually 10–15 minutes or one hour). Thus, durations of activities longer than the threshold are replaced with the maximum allowed duration. In the example in Table 8, the time spent viewing discussion `D4` is exceptionally long, which suggests the likelihood of a long off-task activity. Accounting for these unusually long activities is what we refer to as the **"outlier detection"** problem.

Finally, if a student closes her browser window, then the next time she wants to use the system she is required to log in before she can do anything else. Thus, in some cases, an action is followed by a login action, in which case we know there was certainly some off-task behaviour. The two simple strategies for addressing this issue are 1) to ignore that an action is followed by a login action, if the total duration of the action is less than a given threshold, and 2) to estimate the duration from the remaining records of the given action by a particular user (as done by del Valle and Duffy, 2009). In the example in Table 8, we can see that the time spent viewing resources `R1` and discussions `D5` are certainly overestimated, as they must contain some amount of time spent outside of the system. We refer to this problem as the **"last-action estimation"** problem.

These two problems — outlier detection and last-action estimation — combined with the specifics of Moodle action tracing strategy make time-on-task estimation extremely challenging and require the development of different approaches for time-on-task estimation.

5.2 Experimental Procedure

Given the previously described details of time-on-task estimation and its two main challenges (i.e., "outlier detection" and "last action estimation"), we conducted an experiment using 15 different strategies for time-on-task estimation (Table 9). We selected these particular strategies in order to provide as many different time-on-task estimation strategies as possible. For some of the strategies, we found evidence in the existing literature (Ba-Omar et al., 2007; Grabe & Sigler, 2002; Munk & Drlík, 2011; del Valle & Duffy, 2009; Wise, Zhao, et al., 2013), while others are included in order to provide a comprehensive evaluation of possible time-on-task estimation methods.

The first six strategies completely ignore outlier detection and simply use the actual values from the action logs (this is denoted by x: in their name). However, they differ in how they process the last action of each session. The first strategy (x:x) completely ignores time-on-task estimation challenges and simply calculates the duration of actions by subtracting actual values from the action log (i.e., naïve approach). The second strategy x:ev is similar, except that the duration of the last action of each session is estimated as a mean value of the logs for the same action (e.g., discussion view) of a particular user. On the other hand, the third strategy x:rm estimates the duration of last actions in every session as being 0 seconds. Given that time-on-task estimates are typically used to calculate cumulative time spent on each individual action, this strategy effectively removes a given record from the total sum (as it is estimated being 0 seconds long). Strategies x:l60, x:l30 and x:l10 on the other hand instead of estimating or removing the last action, put an upper value for the duration at 60, 30 and 10 minutes, respectively.

Table 9: Different time-on-task extraction strategies

#	Name	Description
	Group 1:	***No outliers processing, different processing of last actions***
1	x:x	No outliers and last action processing.
2	x:ev	No outliers processing, estimation of last action duration.
3	x:rm	No outliers processing, removal of last action.
4	x:l60	No outliers processing, 60 min last action duration limit.
5	x:l30	No outliers processing, 30 min last action duration limit.
6	x:l10	No outliers processing, 10 min last action duration limit.
	Group 2:	***Thresholding outliers and last actions***
7	l60	60 min duration limit.
8	l30	30 min duration limit.
9	l10	10 min duration limit.
	Group 3:	***Thresholding outliers and estimating last actions***
10	l60:ev	60 min duration limit, last actions estimated.
11	l30:ev	30 min duration limit, last actions estimated.
12	l10:ev	10 min duration limit, last actions estimated.
	Group 4:	***Estimating outliers and last actions***
13	+60ev	Estimate last actions and actions longer than 60 min.
14	+30ev	Estimate last actions and actions longer than 30 min.
15	+10ev	Estimate last actions and actions longer than 10 min.

The second group (l60, l30, and l10) are very simple strategies that put an upper limit on the duration of any action. If an action is shorter, an actual time is used; otherwise, the action is replaced with a particular threshold value. The challenge of this group of strategies is that it is hard to pick a threshold value that would remove as much of the off-task behaviour as possible, while not affecting genuinely long actions.

The third set of strategies (l60:ev, l30:ev, and l10:ev) also place an upper estimate on the duration of all actions, except those followed by a login action (i.e., sessions' last actions). The actions followed by a login action are estimated to be the average duration of a given action, calculated separately for each student. The rationale ascribed here is that if a student performed a particular action many times where it was not followed by a login action, then those records could be used to estimate reasonably accurately the durations for those cases where an action was followed by a login.

Finally, strategies in the last group (+60ev, +30ev, and +10ev) are the most flexible, and they estimate durations of all actions above a particular threshold as an average value for a given action (for a particular user). The rationale is that most actions are very short, and thus actions with extensively long times most likely involve some off-task behaviour, which warrants estimation of their durations based on the remaining records, which are more likely to be genuine.

5.3 Statistical Analysis

In order to examine the level of effect different time-on-task estimation procedures have on the results of different analytical models, we conducted a series of multiple linear regression analyses. There are several reasons for selecting multiple regression models. First, different forms of general linear models — including multiple linear regression — are widely used in diverse research areas (Hastie, Tibshirani, & Friedman, 2013), including learning analytics and EDM (Romero & Ventura, 2010). In addition, multiple linear regression is one of the simplest and most robust models (Hastie et al., 2013) and is one of the methods that should be the least susceptible to changes in time-on-task measures. Finally, given that standardized regression coefficients are easy to interpret and directly comparable, we can easily compare several time-on-task extraction procedures.

6 RESULTS: ONLINE COURSE DATASET

6.1 Overview

A series of multiple regression analyses were undertaken for each of the five performance measures across all 15 time-on-task extraction strategies. Figure 1 shows obtained R^2 values while Table 11 shows the detailed regression results. For all dependent variables, time-on-task measures obtained higher R^2 *v*alues that count measures, which is expected given that they better capture student engagement. What is more interesting is that the differences between estimation strategies are quite substantial. Table 10 shows the summary of the differences between the "worst" and "best" performing strategies. On average, the difference in R^2 was 0.15, which corresponds to 15% of the variance being explained solely by the adoption of a particular time-on-task estimation strategy. The differences were the smallest for the ColHigh measure (R^2 difference of 0.07) and largest for the FinalGrade measure (R^2 difference of 0.23).

Table 10: Summary of differences in R^2 scores between different time-on-task estimation strategies

	R^2				
Performance Measure	***Min***	***Max***	***Range***	***Mean***	***SD***
TMA2Grade	0.08	0.26	0.18	0.14	0.04
TMA3Grade	0.04	0.17	0.12	0.09	0.04
ParticipationGrade	0.23	0.37	0.13	0.3	0.04
FinalGrade	0.06	0.28	0.23	0.16	0.05
ColHigh	0.21	0.28	0.07	0.26	0.02

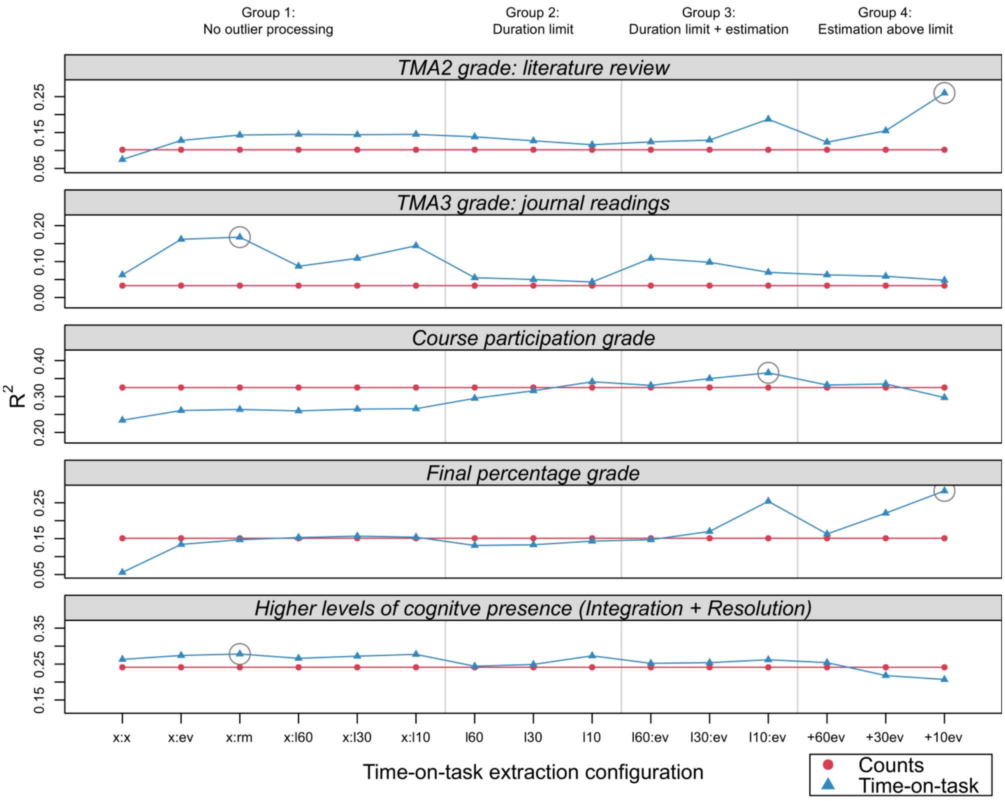


Figure 1: Variation in R2 scores across different time-on-task extraction strategies for five performance measures.

Table 11: Regression results for different time-on-task extraction strategies. Boldface indicates statistical significance at α=.05 level, while gray shade indicates configuration with highest R^2 scores

DV	IV	x:x	x:ev	x:rm	x:l60	x:l30	x:l10	l60	l30	l10	l60:ev	l30:ev	l10:ev	+60ev	+30ev	+10ev
TMA2Grade	*p*-value	0.34	0.07	**0.05**	**0.04**	**0.04**	**0.04**	0.05	0.08	0.11	0.08	0.07	**0.01**	0.09	**0.03**	**0**
	R^2	0.075	0.128	0.143	0.145	0.144	0.145	0.138	0.127	0.116	0.124	0.129	0.187	0.123	0.155	0.26
β coefficients	Assign.ViewTime	0.13	**0.27**	0.27	0.19	0.22	0.25	0.1	0.1	0.1	0.28	**0.31**	**0.34**	**0.28**	**0.3**	**0.27**
	Res.ViewTime	0.05	0.03	0.11	0.15	0.13	0.12	0.19	0.19	0.19	-0.01	-0.09	**-0.31**	-0.1	-0.26	**-0.43**
	Disc.ViewTime	0.02	-0.01	-0.05	0.04	0.01	-0.03	0.07	0.04	-0.01	-0.02	-0.01	0.06	0.01	0.08	0.11
	AddPostTime	-0.05	-0.06	-0.05	-0.08	-0.07	-0.06	-0.14	-0.1	0.02	-0.1	-0.06	0.07	-0.05	0	0.11
	UpdatePostTime	**0.25**	**0.27**	**0.27**	**0.26**	**0.26**	**0.27**	**0.25**	**0.25**	0.22	**0.26**	**0.25**	0.17	0.22	0.2	0.12
TMA3Grade	*p*-value	0.45	**0.03**	**0.02**	0.26	0.14	**0.05**	0.54	0.59	0.67	0.14	0.19	0.39	0.45	0.49	0.61
	R^2	0.063	0.162	0.168	0.087	0.109	0.144	0.055	0.05	0.043	0.109	0.098	0.07	0.063	0.059	0.048
β coefficients	Assign.ViewTime	0.11	**0.28**	**0.31**	0.11	0.18	0.26	-0.02	-0.01	-0.01	0.22	0.21	0.14	0.12	0.14	0.05
	Res.ViewTime	0.11	0.19	0.17	0.15	0.15	0.16	0.14	0.13	0.11	0.14	0.12	0.12	0.08	0.03	0.24
	Disc.ViewTime	0.04	-0.07	-0.09	0.03	-0.01	-0.06	0.06	0.05	0.04	-0.04	-0.03	0	0.02	0.05	0.03
	AddPostTime	-0.07	-0.09	-0.08	-0.1	-0.09	-0.09	-0.04	-0.03	-0.01	-0.04	-0.04	-0.01	0.02	-0.02	0.02
	UpdatePostTime	0.19	**0.23**	**0.23**	0.2	0.21	0.23	0.17	0.17	0.15	0.19	0.19	0.17	0.16	0.16	0.13
Part.Grade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	R^2	0.234	0.261	0.264	0.26	0.265	0.266	0.295	0.316	0.341	0.331	0.35	0.366	0.332	0.335	0.297
β coefficients	Assign.ViewTime	-0.04	-0.01	0	-0.03	-0.02	-0.01	0.01	0.01	0.01	0.11	0.11	0.09	0.1	0.09	0.06
	Res.ViewTime	0.12	0.2	0.18	0.21	0.2	0.19	0.13	0.11	0.11	0.09	0.07	0.08	0.04	0.03	0.13
	Disc.ViewTime	-0.16	0.11	0.12	0.13	0.14	0.13	0.11	0.13	0.13	0.13	0.16	0.17	0.21	0.22	0.2
	AddPostTime	**0.43**	**0.34**	**0.34**	**0.34**	**0.34**	**0.34**	**0.43**	**0.45**	**0.48**	**0.43**	**0.45**	**0.48**	**0.43**	**0.46**	**0.43**
	UpdatePostTime	0.13	0.12	0.12	0.11	0.11	0.12	0.06	0.03	-0.01	0.06	0.02	-0.02	0.03	-0.03	0
FinalGrade	*p*-value	0.49	0.05	**0.03**	**0.03**	**0.02**	**0.03**	0.06	0.05	**0.04**	**0.03**	**0.01**	**0**	**0.02**	**0**	**0**
	R^2	0.056	0.134	0.147	0.153	0.157	0.154	0.131	0.133	0.143	0.147	0.17	0.254	0.163	0.221	0.283
β coefficients	Assign.ViewTime	0.13	0.26	0.24	0.24	0.24	0.24	0.21	0.21	0.23	**0.35**	**0.4**	**0.44**	**0.38**	**0.41**	**0.34**
	Res.ViewTime	0.05	0.03	0.12	0.13	0.13	0.13	0.13	0.13	0.12	-0.06	-0.15	**-0.34**	-0.17	**-0.33**	**-0.43**
	Disc.ViewTime	-0.05	0.08	0.04	0.07	0.06	0.04	0.05	0.05	0	0.03	0.05	0.1	0.08	0.14	0.16
	AddPostTime	0.08	0.03	0.04	0.03	0.03	0.03	-0.01	0	0.1	0	0.03	0.13	0.04	0.08	0.11
	UpdatePostTime	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.16	0.13	0.16	0.14	0.06	0.11	0.09	0.03
ColHigh	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	R^2	0.263	0.274	0.278	0.266	0.272	0.277	0.244	0.249	0.273	0.252	0.254	0.262	0.254	0.218	0.207
β coefficients	Assign.ViewTime	0.02	-0.01	0	-0.05	-0.03	-0.01	-0.09	-0.13	-0.16	-0.15	-0.17	-0.13	-0.21	-0.08	-0.07
	Res.ViewTime	0.12	0.14	0.16	0.17	0.17	0.16	0.13	0.15	0.19	0.06	0.06	0.01	0.05	-0.01	-0.07
	Disc.ViewTime	-0.14	0.11	0.09	0.07	0.08	0.09	0.03	0.04	0.02	0.16	0.17	0.14	0.18	0.12	0.12
	AddPostTime	**0.42**	**0.35**	**0.36**	**0.36**	**0.36**	**0.36**	**0.37**	**0.37**	**0.41**	**0.37**	**0.37**	**0.42**	**0.33**	**0.36**	**0.37**
	UpdatePostTime	**0.22**	0.2	0.2	0.19	0.19	0.2	0.17	0.15	0.12	0.15	0.13	0.09	0.16	0.12	0.11

6.2 Performance Measure Results

6.2.1 TMA2 grade: literature review

For the TMA2 performance measure, all strategies produced higher R^2 values than the count measures, except for the simplest x:x strategy that uses recorded timestamp data without any further adjustments. In terms of R^2 scores, the best performing strategy was +10ev, which estimates the duration of all actions longer than 10 minutes and last session actions as an average of actions recorded for each student. All strategies in the first group (except x:x) and all strategies from the second group achieved similar R^2 scores, while in the third and fourth groups we found the same pattern of increased R^2 with the shortening of the threshold value.

The results of the regression analysis (Table 11) indicate that all models, except the x:x model, were either significant, or marginally non-significant. Still, in terms of the *β* coefficients, there are large differences. For example, the coefficient for time spent updating messages was significant in most of the models from the first three groups, while non-significant in the models in the fourth group. The coefficient for time spent on assignments showed the exact opposite trend. Finally, the coefficient for time spent viewing resources was significant only in two models — including the one with the highest obtained R^2 value, in which the *β* coefficient value was the largest (-0.43).

6.2.2 TMA3 grade: journal readings

For the TMA3 performance measure, all time-on-task estimation strategies gave a better performance than the corresponding count measures. The best performing strategy was the x:rm strategy, which uses recorded timestamp data without any further adjustment, except for the removal of the last action of each session. In general, the strategies from the first and third group achieved better performance than the strategies in the second and fourth group. However, only three regression models from the first group were significant (Table 11). In one of them (x:l10), none of the *β* coefficients were significant, while in the other two models (x:ev and x:rm) the coefficients for the time spent updating messages and viewing assignments were significant, with significantly higher values than in any other model.

6.2.3 Course participation grade

For the ParticipationGrade performance measure, all strategies in the first group obtained R^2 scores lower than the count measures, while other strategies obtained very similar R^2 values as count measures. The highest R^2 score was obtained for the l10:ev strategy, which limits the duration of all actions to 10 minutes, while last session actions were estimated based on other records of the same action for each student.

While all regression models achieved significance (Table 11), there was a large difference between their R^2 values, with the difference of 0.13 between the highest and lowest scoring estimation strategies. Only the regression coefficient for the time spent writing messages was significant in all configurations with its value ranging from 0.34 to 0.48.

6.2.4 Final percentage grade

For the course final percent grade, most time-on-task estimation strategies had scores similar to the count measures. Only the simplest x:x strategy performed significantly worse, while l10, +30ev, and +10ev strategies performed considerably better than the count measures. Similar to the TMA2 performance measure, the highest R^2 scores were obtained with the +10ev strategy.

The detailed regression results shown in Table 11 indicate that four models from the first group and one model from the second group were significant, but without significant *β* coefficients. On the other hand, all models from the third and fourth groups were significant, and all of them had significant regression coefficients for the time spent viewing assignments. The highest scoring model (+10ev) had an R^2 value of

0.28 and significant regression coefficients for the time spent viewing resources (0.–43) and assignments (0.34).

6.2.5 Higher levels of cognitive presence

While the prediction of the count of messages with higher levels of cognitive presence based on time-on-task estimates was better in all but two configurations, the differences were not large. The regression models for all configurations were highly significant, and all of them had a significant regression coefficient only for the time spent posting new messages (Table 11). With the R^2 value of 0.28, the highest performing configuration was x:rm — the same configuration that best predicted TMA2 grades.

7 RESULTS: BLENDED DATASET

Similar to the analysis of a fully online dataset, we conducted a series of multiple linear regression analyses between measures of LMS use and final percent grade for each of the nine courses from the blended dataset. Figure 2 shows the obtained R^2 values, while a more detailed view is given in Table 12. In all but one course (BIOL 1) the best obtained R^2 values were achieved by the use of time-on-task measures. In six courses, the best performing strategy was from the first group (No outlier processing), in two courses, from the second group of strategies (Duration limit), and in one instance (BIOL 1) count measures outperformed all time-on-task estimation strategies.

Regarding the role of time-on-task estimation strategies on the variations in R^2 scores, we observed more modest effects. While in the analyses performed on the online dataset the average range of R^2 was 0.15, in the analyses performed on the blended dataset, we obtained an average range of 0.05 for the R^2 values, indicating that 5% of the variability in the R^2 scores was accounted for solely by a time-on-task estimation strategy. As shown in Figure 2, in the case of the communication (COMM), computer science (COMP), and economics (ECON) courses, the adopted time-on-task estimation strategy had almost zero impact on the obtained R^2 values, and similarly, in the accounting (ACCT) and graphics (GRAP) courses most of the strategies had very similar R^2 values. The largest effect was observed for the two biology courses and for the mathematics course. Interestingly, in case of the first biology (BIOL 1) and the marketing (MARK) courses, count measures outperformed most time-on-task estimation strategies with only the l:10 strategy performing equally as well as the count measures. The biggest benefit from the use of time-on-task measures was achieved for the second biology (BIOL 2) and the mathematics (MATH) courses. With the biology 2 course, the best performing strategies were from the first two groups, while for the mathematics course, the last two groups of strategies performed best.

A closer look at the details of the regression analyses of the blended dataset (Table 13) provides more insight into the observed variations in R^2 scores. In the cases of the ACCT, COMM, COMP, ECON, MARK, and MATH courses, the largest standardized regression coefficients were related to two count measures: the number of Turnitin submissions (TurnitinSubmissionCountLog) and the number of assignment uploads (AssignmentUploadCount). The high predictive power of the two abovementioned count measures were previously reported by several researchers in their analysis of the same dataset (Cho & Kim, 2013; Gašević, Dawson, Rogers, & Gašević, 2015; Trigwell et al., 1999). Given that the used count measures did not change because of the adopted time-on-task estimation strategies and given that they accounted for most of the variability, the effect was very limited. Thus, the use of count measures alongside time-on-task measures limited the effect that different estimation strategies could have on the results of the final regression analyses.

The variations of individual regression coefficients and their significance across different time-on-task estimation strategies show similar variations observed as in the analyses performed on the fully online dataset. In all of the courses, the particular regression coefficients — and more importantly their significance — changed with the time-on-task estimation strategy used. While the use of count measures limited the effect of the adopted time-on-task estimation strategy on the overall predictive power of the model, the latter had a role in shaping the significance levels of different individual predictors — including the count measures.

Table 12: Summary of differences in R^2 scores between different time-on-task estimation strategies

	R^2				
Course	***Min***	***Max***	**Range**	***Mean***	***SD***
ACCT	0.16	0.2	0.04	0.17	0.01
BIOL1	0.12	0.22	0.09	0.17	0.02
BIOL2	0.15	0.26	0.11	0.21	0.04
COMM	0.58	0.6	0.02	0.59	0
COMP	0.53	0.54	0.01	0.54	0
ECON	0.38	0.4	0.02	0.39	0
GRAP	-0.01	0.05	0.06	0.01	0.03
MARK	0.34	0.38	0.03	0.36	0.01
MATH	0.21	0.26	0.06	0.23	0.02

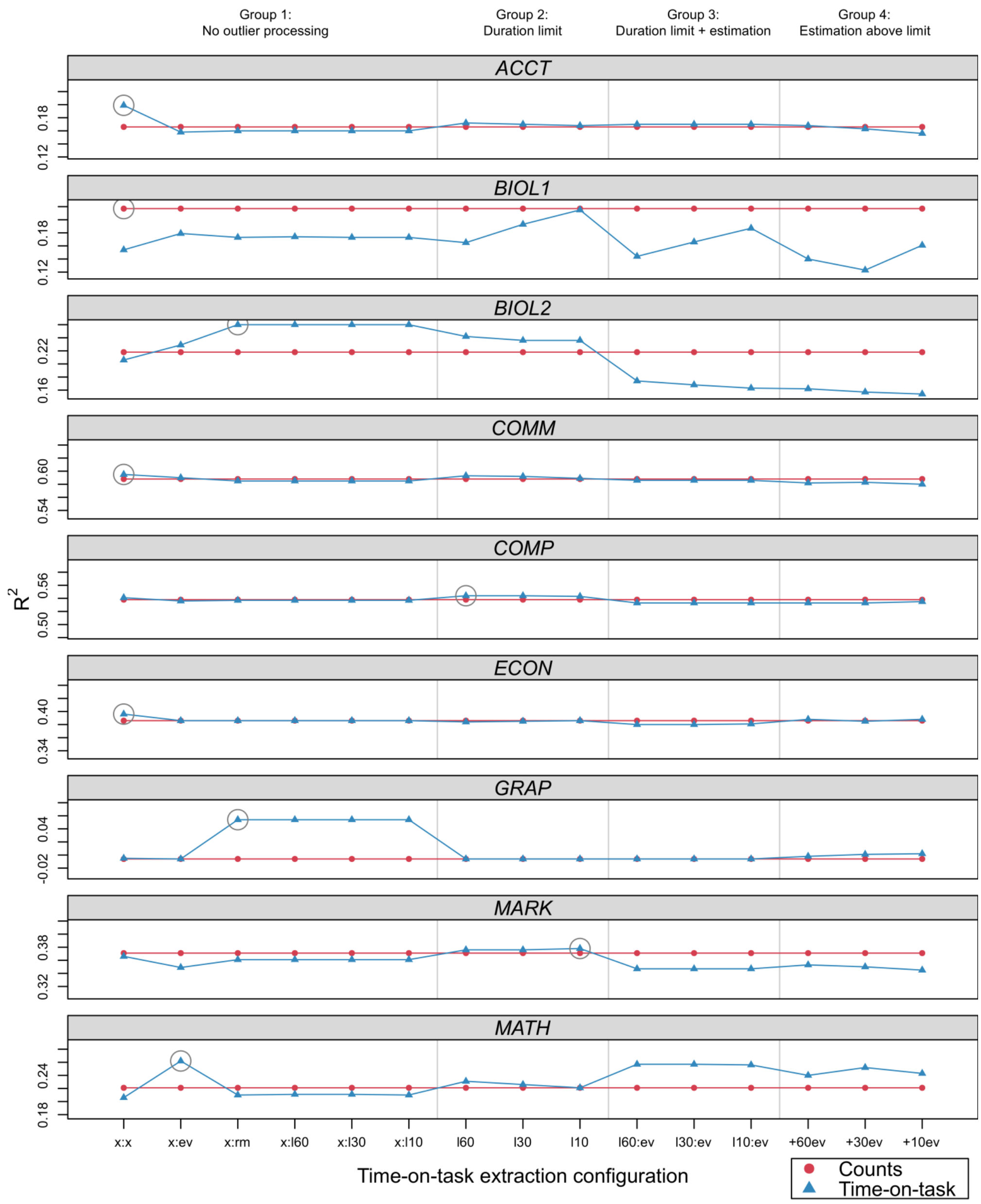


Figure 2: Variation in R^2 scores across different time-on-task extraction strategies for final percentage grade in all nine blended courses.

Table 13: Regression results for different time-on-task extraction strategies. Boldface indicates statistical significance at α=.05 level, while gray shade indicates configuration with highest R^2 scores

DV	IV	x:x	x:ev	x:rm	x:l60	x:l30	x:l10	l60	l30	l10	l60:ev	l30:ev	l10:ev	+60ev	+30ev	+10ev
ACCT FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.199	0.158	0.16	0.16	0.16	0.16	0.172	0.17	0.168	0.17	0.17	0.17	0.168	0.163	0.156
β coefficients	Assign.Upl.Count	**-0.21**	**-0.21**	**-0.19**	**-0.19**	**-0.19**	**-0.19**	**-0.17**	**-0.16**	**-0.16**	**-0.21**	**-0.21**	**-0.21**	**-0.2**	**-0.21**	**-0.21**
	BookPrintCount	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	CourseViewCount	**0.16**	**0.19**	**0.17**	**0.17**	**0.17**	**0.17**	**0.18**	**0.2**	**0.2**	**0.24**	**0.24**	**0.24**	**0.18**	**0.17**	**0.18**
	ForumSearchCount	0.02	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0	0	0	0.01	0.01	0.02
	Turn.Su.CountLog	**0.5**	**0.49**	**0.47**	**0.47**	**0.47**	**0.47**	**0.47**	**0.47**	**0.47**	**0.48**	**0.48**	**0.48**	**0.47**	**0.47**	**0.48**
	Assign.ViewTime	-0.07	0.05	0	0	0	0	-0.05	-0.06	-0.06	0.05	0.05	0.05	0.01	0.01	-0.01
	BookViewTime	**-0.11**	0	**-0.08**	**-0.08**	**-0.08**	**-0.08**	**-0.12**	**-0.11**	**-0.1**	0.02	0.02	0.02	0.01	0.01	-0.02
	ViewDisc.Time	0.03	0.03	-0.01	-0.01	-0.01	-0.01	0.01	0.01	0.01	0.04	0.04	0.04	**0.08**	0.06	0.03
	AddPostTime	0	0	0.02	0.02	0.02	0.02	-0.06	-0.07	-0.07	-0.08	-0.08	-0.08	0.02	0.03	0.02
	GalleryViewTime	-0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.01	-0.01	0.02	0.02	0.02	0.04	0.03	0.01
	Res.ViewTime	**0.16**	-0.04	0.04	0.04	0.04	0.04	**0.11**	**0.1**	**0.1**	**-0.09**	**-0.09**	**-0.09**	**-0.09**	**-0.08**	0
BIOL1 FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.154	0.179	0.173	0.174	0.173	0.173	0.165	0.193	0.215	0.144	0.166	0.187	0.14	0.123	0.161
β coefficients	CourseViewCount	**0.37**	**0.18**	0.15	0.15	0.15	0.15	**0.24**	**0.23**	**0.21**	**0.39**	**0.39**	**0.39**	**0.4**	**0.38**	**0.35**
	ForumSearchCount	0.01	0	0.01	0.01	0.01	0.01	0.03	0.03	0.02	0.03	0.03	0.01	0.03	0.02	0.02
	Assign.ViewTime	0	-0.02	0.02	0.02	0.02	0.02	-0.08	-0.08	-0.03	-0.07	-0.04	-0.02	-0.07	-0.06	**-0.16**
	ViewDisc.Time	**0.15**	0.14	**0.16**	**0.16**	**0.16**	**0.16**	**0.23**	**0.24**	**0.26**	0.01	0	0.01	-0.06	-0.01	-0.07
	AddPostTime	0.03	0.06	0.05	0.05	0.05	0.05	-0.05	-0.07	-0.09	0.02	0	-0.02	0.02	0.05	0.05
	QuizViewTime	0	0.03	0.03	0.03	0.03	0.03	-0.17	**-0.25**	**-0.25**	**-0.2**	**-0.26**	**-0.24**	-0.14	0.02	-0.1
	QuizAttemptTime	0.06	-0.06	-0.04	-0.04	-0.04	-0.04	0.16	**0.29**	**0.35**	0.13	**0.27**	**0.34**	0.09	-0.01	0.04
	QuizReviewTime	0.03	0.09	0.08	0.08	0.08	0.08	0.04	0.05	0.05	0.05	0.04	0.04	0.02	-0.03	-0.02
	Res.ViewTime	0.11	**0.23**	**0.21**	**0.21**	**0.21**	**0.21**	0.12	0.11	0.07	0.07	0.05	0.02	0.03	0.03	0.08
BIOL2 FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.206	0.229	0.26	0.26	0.26	0.26	0.242	0.236	0.236	0.174	0.168	0.163	0.162	0.157	0.154
β coefficients	BookPrintCount	-0.01	0	0	0	0	0	-0.01	-0.01	-0.01	0	0	0	0	0	0
	CourseViewCount	**0.28**	0.05	0.02	0.02	0.02	0.02	0.01	0.01	-0.01	**0.29**	**0.31**	**0.31**	**0.27**	**0.28**	**0.27**
	FeedbackCount	**0.17**	**0.17**	**0.17**	**0.17**	**0.17**	**0.17**	**0.16**	**0.16**	**0.16**	**0.2**	**0.21**	**0.21**	**0.18**	**0.19**	**0.18**
	ForumSearchCount	-0.06	-0.06	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.05	-0.05	**-0.07**	-0.07	**-0.08**
	BookViewTime	0.02	0.04	**-0.07**	**-0.07**	**-0.07**	**-0.07**	-0.04	-0.03	-0.02	0.04	0.04	0.04	0.04	0.04	0.04
	ViewDisc.Time	**-0.08**	-0.05	**-0.1**	**-0.1**	**-0.1**	**-0.1**	-0.07	-0.04	0	0.03	0.03	0.03	0.03	0.03	0.01
	AddPostTime	0.04	0.02	0.04	0.04	0.04	0.04	-0.01	-0.03	-0.04	**-0.11**	**-0.11**	**-0.11**	0.01	-0.01	0.02
	MapViewTime	0.02	0.04	0.02	0.02	0.02	0.02	-0.04	-0.04	-0.06	0.04	0.04	0.04	0.03	0.03	0.03
	QuizViewTime	**0.11**	**0.09**	0.07	**0.07**	**0.07**	0.07	0.05	0.05	0.12	0.06	0.06	0.05	0.05	0.05	0.06
	QuizAttemptTime	0	**0.07**	**0.17**	**0.17**	**0.17**	**0.17**	**0.13**	0.14	0.08	0.02	0.01	0	0.02	0.02	0.02
	QuizReviewTime	0.07	0.05	-0.04	-0.04	-0.04	-0.04	X	X	X	**0.13**	**0.1**	0.06	**0.1**	0.07	0.04
	Res.ViewTime	**0.19**	**0.32**	**0.35**	**0.35**	**0.35**	**0.35**	**0.35**	**0.33**	**0.33**	**-0.07**	**-0.08**	**-0.08**	**-0.08**	**-0.08**	**-0.08**
	AdobeCo.ViewTime	0.02	0	-0.01	-0.01	-0.01	-0.01	0.03	0.02	0.01	0.02	0.02	0.02	0.02	0.03	0
COMM FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.595	0.59	0.585	0.585	0.585	0.585	0.593	0.592	0.589	0.586	0.586	0.586	0.582	0.583	0.58
β coefficients	Assign.Upl.Count	**-0.53**	**-0.58**	**-0.58**	**-0.58**	**-0.58**	**-0.58**	**-0.58**	**-0.58**	**-0.58**	**-0.57**	**-0.57**	**-0.57**	**-0.57**	**-0.56**	**-0.56**
	CourseViewCount	**0.08**	0.05	0.06	0.06	0.06	0.06	0.02	0.03	0.04	**0.09**	**0.09**	**0.09**	**0.12**	**0.12**	**0.12**
	ForumSearchCount	-0.01	-0.01	-0.02	-0.02	-0.02	-0.02	-0.01	-0.01	-0.01	0	0	0	-0.01	0	-0.01
	Turn.Su.CountLog	**1.05**	**1.12**	**1.12**	**1.12**	**1.12**	**1.12**	**1.1**	**1.1**	**1.1**	**1.14**	**1.14**	**1.14**	**1.14**	**1.13**	**1.11**
	Assign.ViewTime	**0.09**	**0.1**	**0.09**	**0.09**	**0.09**	**0.09**	**0.11**	**0.11**	**0.09**	0.02	0.02	0.02	0.02	-0.01	0.03
	ViewDisc.Time	**0.12**	0.03	0.03	0.03	0.03	0.03	0.07	0.06	0.05	0.02	0.02	0.02	0.01	0.01	0.02
	AddPostTime	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01	0.05	0.05	0.05	0.06	0.06	0.06	-0.02	-0.02	-0.01
	Res.ViewTime	0.01	0.05	-0.02	-0.02	-0.02	-0.02	-0.04	-0.04	-0.04	**0.06**	**0.06**	**0.06**	**0.06**	**0.06**	0.01

Table 13 (continued): Regression results for different time-on-task extraction strategies. Boldface indicates statistical significance at α=.05 level, while gray shade indicates configuration with highest R^2 scores

DV	IV	x:x	x:ev	x:rm	x:l60	x:l30	x:l10	l60	l30	l10	l60:ev	l30:ev	l10:ev	+60ev	+30ev	+10ev
COMP FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.541	0.536	0.537	0.537	0.537	0.537	0.544	0.544	0.543	0.533	0.533	0.533	0.533	0.533	0.535
β coefficients	Assign.Upl.Count	**-0.45**	**-0.47**	**-0.47**	**-0.47**	**-0.47**	**-0.47**	**-0.44**	**-0.43**	**-0.43**	**-0.46**	**-0.46**	**-0.46**	**-0.47**	**-0.46**	**-0.46**
	CourseViewCount	**0.13**	**0.12**	**0.13**	**0.13**	**0.13**	**0.13**	**0.14**	**0.14**	**0.14**	**0.13**	**0.13**	**0.13**	**0.13**	**0.13**	**0.13**
	Turn.Su.CountLog	**1.03**	**1.04**	**1.03**	**1.03**	**1.03**	**1.03**	**1.04**	**1.04**	**1.04**	**1.04**	**1.04**	**1.04**	**1.04**	**1.03**	**1.03**
	Assign.ViewTime	-0.03	0.02	0.02	0.02	0.02	0.02	-0.04	-0.04	-0.04	0.02	0.02	0.02	0.03	0	0.01
	QuizViewTime	**-0.1**	-0.04	-0.05	-0.05	-0.05	-0.05	**-0.17**	**-0.2**	-0.2	-0.03	-0.03	-0.03	-0.03	-0.03	-0.05
	QuizAttemptTime	0.01	0.01	-0.02	-0.02	-0.02	-0.02	0.1	0.12	0.1	0.02	0.01	0.02	0.01	0.01	0.02
	QuizReviewTime	0.01	-0.05	-0.04	-0.04	-0.04	-0.04	0.04	0.06	0.07	-0.01	0	0	-0.01	-0.02	-0.04
	Res.ViewTime	0.04	0	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0	0	0	0.01	0.01	0
ECON FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.396	0.386	0.386	0.386	0.386	0.386	0.384	0.385	0.386	0.38	0.38	0.381	0.388	0.385	0.388
β coefficients	Assign.Upl.Count	**-0.43**	**-0.45**	**-0.45**	**-0.45**	**-0.45**	**-0.45**	**-0.44**	**-0.43**	**-0.42**	**-0.45**	**-0.45**	**-0.45**	**-0.44**	**-0.44**	**-0.45**
	BookPrintCount	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	CourseViewCount	**0.14**	0.08	0.05	0.05	0.05	0.05	0.08	0.1	0.11	**0.13**	**0.13**	**0.12**	**0.17**	**0.17**	**0.16**
	ForumSearchCount	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	Turn.Su.CountLog	**0.86**	**0.87**	**0.88**	**0.88**	**0.88**	**0.88**	**0.88**	**0.89**	**0.88**	**0.87**	**0.87**	**0.87**	**0.84**	**0.85**	**0.86**
	Assign.ViewTime	0.01	0	-0.01	-0.01	-0.01	-0.01	-0.07	**-0.09**	**-0.11**	-0.01	-0.01	-0.01	-0.06	**-0.06**	-0.05
	BookViewTime	0	-0.01	-0.02	-0.02	-0.02	-0.02	0.03	0.03	0.02	-0.01	-0.01	-0.01	-0.02	-0.01	0.03
	ViewDisc.Time	0.06	0.02	0.04	0.04	0.04	0.04	0.02	0.01	0	0	0	0	0	0.01	0.02
	AddPostTime	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.03	0.02	0.05
	QuizViewTime	-0.02	0	0.05	0.05	0.05	0.05	0	-0.01	0.01	-0.03	-0.03	-0.03	-0.04	-0.04	-0.07
	QuizAttemptTime	-0.01	-0.01	0.05	0.05	0.05	0.05	0.13	0.14	0.13	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01
	QuizReviewTime	0.04	0.05	-0.01	-0.01	-0.01	-0.01	-0.1	-0.1	-0.11	0.04	0.04	0.04	0.05	0.06	0.06
	Res.ViewTime	**0.12**	**0.12**	**0.09**	**0.09**	**0.09**	**0.09**	0.08	0.07	0.09	0.04	0.04	0.06	**-0.06**	-0.01	-0.02
GRAP FinalGrade	*p*-value	0.56	0.64	**0**	**0**	**0**	**0**	0.62	0.64	0.61	0.64	0.64	0.64	0.42	0.35	0.32
	adj. R^2	-0.005	-0.006	0.054	0.054	0.054	0.054	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006	-0.002	0.001	0.002
β coefficients	CourseViewCount	0.07	0.07	**0.18**	**0.18**	**0.18**	**0.18**	0.09	0.07	0.05	0.07	0.07	0.07	0.08	0.08	0.09
	Res.ViewTime	0.04	-0.01	**-0.27**	**-0.27**	**-0.27**	**-0.27**	-0.03	0.01	0.03	0	0	0	0.07	0.08	0.09
MARK FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.366	0.349	0.361	0.361	0.361	0.361	0.376	0.376	0.378	0.347	0.347	0.347	0.353	0.35	0.345
β coefficients	Assign.Upl.Count	**-0.45**	**-0.48**	**-0.46**	**-0.46**	**-0.46**	**-0.46**	**-0.45**	**-0.44**	**-0.44**	**-0.47**	**-0.47**	**-0.47**	**-0.46**	**-0.46**	**-0.47**
	CourseViewCount	**0.14**	**0.15**	**0.15**	**0.14**	**0.15**	**0.15**	**0.2**	**0.23**	**0.26**	**0.18**	**0.18**	**0.18**	**0.16**	**0.16**	**0.16**
	ForumSearchCount	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	Turn.Su.CountLog	**0.88**	**0.89**	**0.88**	**0.88**	**0.88**	**0.88**	**0.87**	**0.88**	**0.88**	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.88**
	Assign.ViewTime	**-0.08**	0.02	-0.01	0	-0.01	-0.01	-0.04	-0.06	**-0.08**	0.01	0.01	0.01	**0.06**	0.01	-0.01
	ChatViewTime	0	0	0	0	0	0	-0.03	-0.02	-0.01	0	0	0	0	0	0
	ChatTalkTime	-0.04	-0.03	-0.02	-0.02	-0.02	-0.02	0.01	0.01	0.01	-0.02	-0.02	-0.02	**-0.07**	**-0.07**	-0.03
	ViewDisc.Time	0.03	-0.04	-0.08	-0.08	-0.08	-0.08	**-0.18**	**-0.19**	**-0.21**	-0.02	-0.02	-0.02	-0.03	-0.03	0.01
	AddPostTime	-0.05	-0.04	-0.03	-0.03	-0.03	-0.03	0.01	0	0	-0.06	-0.07	**-0.07**	-0.03	-0.04	-0.04
	Res.ViewTime	**0.11**	0.06	**0.13**	**0.13**	**0.13**	**0.13**	**0.17**	**0.15**	**0.14**	-0.03	-0.03	-0.03	-0.02	-0.03	-0.01
MATH FinalGrade	*p*-value	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	adj. R^2	0.206	0.262	0.21	0.211	0.211	0.21	0.231	0.226	0.221	0.257	0.257	0.256	0.24	0.252	0.243
β coefficients	Assign.Upl.Count	**-0.46**	**-0.45**	**-0.45**	**-0.45**	**-0.45**	**-0.45**	**-0.45**	**-0.45**	**-0.46**	**-0.49**	**-0.48**	**-0.49**	**-0.45**	**-0.42**	**-0.41**
	CourseViewCount	**0.33**	**0.2**	**0.22**	**0.22**	**0.22**	**0.22**	0.06	0.1	0.14	**0.25**	**0.26**	**0.27**	**0.36**	**0.34**	**0.32**
	ForumSearchCount	0.01	0.01	0.01	0.01	0.01	0.01	0	0	-0.01	-0.02	-0.02	-0.02	0	0.01	0.01
	Turn.Su.CountLog	**0.64**	**0.65**	**0.63**	**0.63**	**0.63**	**0.63**	**0.58**	**0.59**	**0.6**	**0.66**	**0.66**	**0.66**	**0.65**	**0.61**	**0.6**
	Assign.ViewTime	-0.05	**-0.11**	-0.01	-0.01	-0.01	-0.01	0.1	0.09	0.06	-0.11	-0.11	-0.11	X	-0.02	-0.03
	ViewDisc.Time	0.08	**0.19**	0.08	0.09	0.09	0.08	0.1	0.1	0.1	**0.17**	**0.17**	**0.17**	**0.14**	**0.14**	0.04
	AddPostTime	-0.06	-0.05	-0.05	-0.05	-0.05	-0.05	0.09	0.1	0.1	0.12	0.12	0.12	-0.06	-0.06	-0.06
	Res.ViewTime	0.02	**0.18**	0.13	0.13	0.13	0.13	0.16	0.13	0.1	0.08	0.08	0.06	-0.11	**-0.18**	**-0.19**

8 DISCUSSION

8.1 Discussion of the Results with the Online Course Dataset

From the results of multiple regression models, investigating the effect of different time-on-task estimation strategies on five different performance measures, we can confirm that the *choice of a particular time-on-task estimation strategy plays an important role in the overall model fit and subsequent model interpretation*. The average R^2 range of 0.15 implies that a large proportion of variability can be explained solely by the adopted estimation strategy. Even more importantly, the significance of the overall model, its β coefficients, and their statistical significance were not consistent for three of the five models (i.e., TMA2 grade, TMA3 grade, and final grade) indicating the important role of the adopted time-on-task estimation strategy on the analysis results and conclusions that can be drawn from these results. However, we cannot say whether the higher scoring models are overfitting the data (i.e., type I error), or that the lower scoring models do not properly fit the data (i.e., type II error). The answer to this question depends on the availability of field observational data and this is a suggested direction for future work.

The comparison of the different estimation strategies across the five performance measures indicated that not a single measure was a clear “winner.” Simply put, the results did not reveal a measure that outperformed all other strategies for all dependent variables. Different strategies provided the best fit for the five selected performance measures. Interestingly, the first group of strategies, which generally allows for a much longer duration of action than other strategies, performed worse than count measures for predicting course participation grade, and better for predicting the TMA2 grade, TMA3 grade, and the number of messages with higher levels of cognitive presence (CoIHigh). As the participation grade was not given based on the total time spent on discussions, but rather based on students’ observable behaviour (i.e., active engagement via message posting), the count measures provided a better fit to the data, especially when compared to the first group of strategies that ignored the issues of student off-task behaviour. For measures more related to the quality of student output — such as the TMA2 grade, the TMA3 grade, and the number of messages with higher levels of cognitive presence — the estimation strategies in the first group provided a better fit for the data, as they inherently better captured the total amount of effort that students invested.

If we move the discussion from individual strategies to groups of strategies, we can see that the only group that consistently outperformed the count measures was the third group of strategies. The third group put a particular upper limit on the duration of all actions and estimated the durations of last session actions based on other recordings of the action in question for each student. However, more research using observational data is required to answer conclusively whether those estimation strategies are indeed the most accurate ones.

8.2 Discussion of the Results with the Blended Courses Dataset

One of the goals of the analyses performed with the blended dataset was to examine further on a larger dataset the observed effect of different time-on-task estimation strategies. The results of the second set of the multiple regression analyses provided a further confirmation that time-on-task estimation strategy plays an important role in shaping the final results of statistical analyses. The overall R^2 values*,* alongside individual regression coefficients and their statistical significance, were varied considerably across different time-on-task estimation strategies. However, in contrast to the first experiment where the average variation in R^2 was 0.15, the average variation of R^2 values in the range of 0.05 for the blended dataset implies that inclusion of count measures can lower the effect of the adopted time-on-task estimation strategy on the overall predictive power of the statistical model. These results were not completely unexpected, as inclusion of count or any other measures lowers the relative contribution of time-on-task measures to the overall model fit, which in turn produces less variation across different time-on-task estimation strategies. This is particularly evident in models where certain count measures — such as the number of turnitin submissions — have a strong predictive power themselves and thus remove the overall significance of extracted time-on-task measures.

The comparison of different time-on-task estimation strategies across different courses in the blended dataset — similarly to the results from the online dataset — reveals that not a single time-on-task estimation strategy was the clear winner. In many courses (i.e., ACCT, BIOL 2, COMM, ECON, GRAPH, and MATH), the first group of strategies that enabled longer action durations provided a better fit than those of time-on-task estimation. While in other courses (i.e., COMP and MARK), the second group of strategies provided better results. Interestingly, the last two groups of estimation strategies — those that provided the best fit in three out of the five cases in the analyses of the online dataset — were not the best performing in any course. Only in the case of the mathematics course, the third and fourth group of strategies provided similar results as the best performing x:rm strategy from the first group. The investigation about the underlying reasons for the observed differences between the findings of the analyses of both datasets provide an important direction for further research.

8.3 General Discussion

Comparing the results of the analyses of the two datasets (Figure 1 and Figure 2) indicates that only count measures provided a reasonably good fit for the blended dataset. For the online dataset, the estimation of all the performance measures — except participation grade — benefited substantially from using time-on-task measures, almost regardless of the adopted estimation strategy. In the analyses of the blended dataset, however, the count measures provided a better fit than most of the time-on-task measures. Given that the course in the online dataset was a fully online distance education course and that all nine courses in the blended dataset were blended courses, the relative amount of activity per student is much higher in the fully online course. The fully online course had a much higher volume of student activity than the blended courses, as seen in the comparison of the values shown in Table 1 and Table 4. On average, each session of the fully online course had about four times more actions and over 20 times more messages than each of the blended courses in the second dataset. Given this clear difference in the two datasets, it is very likely that the importance of time-on-task estimation is more critical for fully online courses that depend almost entirely on online learning systems for any form of interaction between students, instructors, and content. Although this seems likely, it warrants further investigation and would be one of the directions for further research.

8.4 Implications for the Learning Analytics Community

Several practical implications arise from the results of the present study. Above all is the need for more caution when using time-on-task measures for building learning analytics models. Given that details of time-on-task estimation can potentially impact reported research findings, appropriately addressing time-on-task estimation becomes a critical part of standard research practice in the learning analytics community. This is particularly true in cases where time-on-task measures are not accompanied by additional measures such as counts of relevant activities.

Another important implication of this paper is that perhaps the role of time-on-task in learning analytics research should be reconsidered. With all the challenges in accurate estimation of time-on-task, given the off-task behaviours, and without a methodologically clear estimation strategy, perhaps using time-on-task measures should be reconsidered and counts measures be more promoted. This is particularly true given the need for more replication studies in the learning analytics field and for clear, sound, easily reported, replicable data-analysis strategies. Evidence of the benefits of time-on-task measures on the final model performance exists, but the question is whether those benefits outweigh the methodological and practical disadvantages associated with their use.

As Karweit (1984) urged educational researchers of the 1980s to pay attention to the challenges of time-on-task estimation in traditional classrooms, so too do we want to draw the attention of the present day global learning analytics community to the same issue. Given that modern technology provides many opportunities for multi-tasking and distractions (e.g., Calderwood et al., 2014; Judd, 2014; Rosen et al., 2013), we strongly argue that time-on-task estimation, its issues, limits, and reliability challenges warrant further consideration.

8.5 Limitations

The primary limitation of this study is related to our inability to generalize from the presented results and decisively point to the overall “best” method for time-on-task estimation. The performance of different estimation strategies depends on the particular characteristics of the target course. Given that we do not have observational field data that would provide accurate measures for students’ actual time-on-task, it is currently not possible to give conclusive recommendations for selection of time-on-task estimation strategies. Furthermore, the present study examined only the effects of time-on-task measuring procedures on one particular statistical model (i.e., multiple linear regression), and it is likely that this also plays a role in shaping the results of the present study.

8.6 Future Work

While this study provides insights into the effects of different time-on-task estimation methods on the results of several analytical models, there are some potential areas for improvement and future work. First, similar to the work done by Baker (2007), Cetintas et al. (2009), Cetintas et al. (2010), Roberge, Rojas, and Baker (2012), and Judd (2014), it would be very helpful to gather "gold standard" data — accurate empirical data about student time-on-task — that could be used to 1) define best practices in time-on-task estimation, and 2) develop automated tools for time-on-task extraction and detection of off-task behaviour. Second, the current study only investigated the effects of different time-on-task estimation strategies on the results of multiple regression models. It would be interesting to see the effects on other types of models; for example, classification systems for automated student grading. Third, the analysis of the observed differences between online and blended courses is important to examine to what extent the particular form of delivery moderates the effects of time-on-task estimation. Finally, it the spirit of open and reproducible research, it would be very useful — from a practical perspective — to develop a standardized plugin for the extraction of trace data from popular LMS systems (e.g., Moodle, WebCT, Sakai, Canvas) that could provide fast and easy-to-use access to time-on-task and count measures.

9 CONCLUSIONS

In this paper, we presented a study that looked at the different approaches for estimating students' time-on-task behaviour based on LMS trace data. We examined 15 different time-on-task estimation strategies and investigated the consequences of adopting various estimation approaches on the results of five learning analytics models of student performance. We also compared time-on-task and count measures in terms of how well they explain the student differences in the five performance measures. Our results indicate that, for the most part, time-on-task estimates outperform count data. However, the adoption of a particular time-on-task estimation strategy can have a significant effect on the overall fit of the model, its significance, and eventually on the interpretation of research findings. With the rising amount of student distraction by digital technology, researchers should be aware of the role that noise in the LMS trace data can play on developed analytics.

There are several important consequences of the presented study. First, the learning analytics community should recognize the importance of time-on-task estimation and the role it plays in the quality of analytical models and their interpretation. Second, with the goal of providing better groundwork for open, replicable, and reproducible research, published literature should address the time-on-task estimation process in sufficient detail. Finally, with the goal of providing a set of standards and common practices for conducting learning analytics research, this paper calls for further investigation of the issues related to student time-on-task estimation.

REFERENCES

Admiraal, W., Wubbels, T., & Pilot, A. (1999). College teaching in legal education: Teaching method, students' time-on-task, and achievement. *Research in Higher Education*, *40*(6), 687–704. http://dx.doi.org/10.1023/A:1018712914619

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '04), 383–390. http://dx.doi.org/10.1145/985692.985741

Baker, R. S. J. d. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07), 1059–1068. http://dx.doi.org/10.1145/1240624.1240785

Ba-Omar, H., Petrounias, I., & Anwar, F. (2007). A framework for using web usage mining to personalise e-learning. *Proceedings of the 7th IEEE International Conference on Advanced Learning Technologies* (ICALT '07), 937–938. http://dx.doi.org/10.1109/ICALT.2007.13

Bloom, B. S. (1974). Time and learning. *American Psychologist*, *29*(9), 682–688. http://dx.doi.org/10.1037/h0037632

Bowman, L. L., Waite, B. M., & Levine, L. E. (2015). Multitasking and attention. In L. D. Rosen, N. A. Cheever, & L. M. Carrier (Eds.), *The Wiley Handbook of Psychology, Technology, and Society* (pp. 388–403). John Wiley & Sons, Ltd. http://dx.doi.org/10.1002/9781118771952.ch22

Brown, A. H., & Green, T. (2009). Time students spend reading threaded discussions in online graduate courses requiring asynchronous participation. *The International Review of Research in Open and Distance Learning*, *10*(6), 51–64. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/760/1432

Calderwood, C., Ackerman, P. L., & Conklin, E. M. (2014). What else do college students "do" while studying? An investigation of multitasking. *Computers & Education*, *75*, 19–29. http://dx.doi.org/10.1016/j.compedu.2014.02.004

Carroll, J. (1963). A model of school learning. *Teachers College Record*, *64*(8), 723–733.

Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the world-wide web. *Proceedings of the 3rd International World-Wide Web Conference: Technology, Tools and Applications* (WWW3), 1065–1073. http://dx.doi.org/10.1016/0169-7552(95)00043-7

Cetintas, S., Si, L., Xin, Y. P., & Hord, C. (2010). Automatic detection of off-task behaviors in intelligent tutoring systems with machine learning techniques. *IEEE Transactions on Learning Technologies*, *3*(3), 228–236. http://dx.doi.org/10.1109/TLT.2009.44

Cetintas, S., Si, L., Xin, Y. P., Hord, C., & Zhang, D. (2009). Learning to identify students' off-task behavior in intelligent tutoring systems. In V. Dimitrova, R. Mizoguchi, B. du Boulay, A. Graesser (Eds.), *Building learning systems that care: Proceedings of the 14th International Conference on Artificial Intelligence in Education* (AIED '09), (pp. 701–703). Amsterdam, the Netherlands:IOS Press.

Chickering, A. W., & Gamson, Z. F. (1989). Seven principles for good practice in undergraduate education. *Biochemical Education*, *17*(3), 140–141. http://dx.doi.org/10.1016/0307-4412(89)90094-0

Chitraa, V., & Davamani, A. S. (2010). A survey on preprocessing methods for web usage data. *International Journal of Computer Science and Information Security*, *7*(3), 78–83.

Cho, M. H., & Kim, B. J. (2013). Students' self-regulation for interaction with others in online learning environments. *The Internet and Higher Education*, *17*, 69–75. http://dx.doi.org/10.1016/j.iheduc.2012.11.001

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th edition). London/New York: Routledge.

Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: Information and pattern discovery on the world wide web. *Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence* (ICTAI '97), 558–567. http://dx.doi.org/10.1109/TAI.1997.632303

Cooley, R., Mobasher, B., & Srivastava, J. (1999). Data preparation for mining world wide web browsing patterns. *Knowledge and Information Systems*, *1*(1), 5–32. http://dx.doi.org/10.1007/BF03325089

del Valle, R., & Duffy, T. M. (2009). Online learning: Learner characteristics and their approaches to managing learning. *Instructional Science*, *37*(2), 129–149. http://dx.doi.org/10.1007/s11251-007-9039-0

García-Martín, J., & García-Sánchez, J.-N. (2013). Patterns of Web 2.0 tool use among young Spanish people. *Computers & Education*, *67*, 105–120. http://dx.doi.org/10.1016/j.compedu.2013.03.003

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2–3), 87–105. http://dx.doi.org/10.1016/S1096-7516(00)00016-6

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23. http://dx.doi.org/10.1080/08923640109527071

Gašević, D., Adesope, O., Joksimović, S., & Kovanović, V. (2015). Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, *24*, 53–65. http://dx.doi.org/10.1016/j.iheduc.2014.09.006

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, 28*, 68–84. http://dx.doi.org/10.1016/j.iheduc.2015.10.002

Grabe, M., & Sigler, E. (2002). Studying online: Evaluation of an online study environment. *Computers & Education*, *38*(4), 375–383. http://dx.doi.org/10.1016/S0360-1315(02)00020-9

Guo, Y., Wang, J., Moore, J., Liu, M., & Chen, H.-L. (2009). A case study of usability testing on an asynchronous e-learning platform. *Proceedings of the 2009 Joint Conferences on Pervasive Computing (JCPC),* 693–698. http://dx.doi.org/10.1109/JCPC.2009.5420093

Hastie, T. J., Tibshirani, R. J., & Friedman, J. H. (2013). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Helmke, A., Schneider, W., & Weinert, F. E. (1986). Quality of instruction and classroom learning outcomes: The German contribution to the IEA classroom environment study. *Teaching and Teacher Education*, *2*(1), 1–18. http://dx.doi.org/10.1016/0742-051X(86)90002-8

Hewitt, J., Brett, C., & Peters, V. (2007). Scan rate: A new metric for the analysis of reading behaviors in asynchronous computer conferencing environments. *American Journal of Distance Education*, *21*(4), 215–231. http://dx.doi.org/10.1080/08923640701595373

Hsu, Y.-C., & Ching, Y.-H. (2013). Mobile app design for teaching and learning: Educators' experiences in an online graduate course. *The International Review of Research in Open and Distance Learning*, *14*(4). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1542

Hussain, T., Asghar, S., & Masood, N. (2010). Web usage mining: A survey on preprocessing of web log file. *Proceedings of the 2010 International Conference on Information and Emerging Technologies* (ICIET 2010), 1–6. http://dx.doi.org/10.1109/ICIET.2010.5625730

Judd, T. (2014). Making sense of multitasking: The role of Facebook. *Computers & Education*, *70*, 194–202. http://dx.doi.org/10.1016/j.compedu.2013.08.013

Karweit, N. (1984). Time-on-task reconsidered: Synthesis of research on time and learning. *Educational Leadership*, *41*(8), 32–35.

Karweit, N., & Slavin, R. E. (1982). Time-on-task: Issues of timing, sampling, and definition. *Journal of Educational Psychology*, *74*(6), 844–851. http://dx.doi.org/10.1037/0022-0663.74.6.844

Kolloffel, B., Eysink, T. H. S., & de Jong, T. (2011). Comparing the effects of representational tools in collaborative and individual inquiry learning. *International Journal of Computer-Supported Collaborative Learning*, *6*(2), 223–251. http://dx.doi.org/10.1007/s11412-011-9110-3

Kraus, L. A., Reed, W. M., & Fitzgerald, G. E. (2001). The effects of learning style and hypermedia prior experience on behavioral disorders knowledge and time on task: A case-based hypermedia environment. *Computers in Human Behavior*, *17*(1), 125–140. http://dx.doi.org/10.1016/S0747-5632(00)00030-3

Lust, G., Elen, J., & Clarebout, G. (2013a). Regulation of tool-use within a blended course: Student differences and performance effects. *Computers & Education*, *60*(1), 385–395. http://dx.doi.org/10.1016/j.compedu.2012.09.001

Lust, G., Elen, J., & Clarebout, G. (2013b). Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. *Computers in Human Behavior*, *29*(5), 2013–2021. http://dx.doi.org/10.1016/j.chb.2013.03.014

Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., & Clarebout, G. (2011). Tool-use in a blended undergraduate course: In search of user profiles. *Computers & Education*, *57*(3), 2135–2144. http://dx.doi.org/10.1016/j.compedu.2011.05.010

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599. http://dx.doi.org/10.1016/j.compedu.2009.09.008

Marquardt, C. G., Becker, K., & Ruiz, D. D. (2004). A pre-processing tool for web usage mining in the distance education domain. *Proceedings of the International Database Engineering and Applications Symposium* (IDEAS '04), 78–87. http://dx.doi.org/10.1109/IDEAS.2004.2

Mobasher, B., Cooley, R., & Srivastava, J. (1999). Creating adaptive web sites through usage-based clustering of URLs. In P. Scheuermann (Ed.), *Proceedings of the1999 Workshop on Knowledge and Data Engineering Exchange* (KDEX '99), (pp. 19–25). Los Alamitos: IEEE Computer Society.

Moodle HQ. (2014). [Computer Software]. Retrieved from https://moodle.org/

Munk, M., & Drlík, M. (2011). Impact of different pre-processing tasks on effective identification of users' behavioral patterns in web-based educational system. *Procedia Computer Science*, *4*, 1640–1649. http://dx.doi.org/10.1016/j.procs.2011.04.177

Munk, M., Kapusta, J., & Švec, P. (2010). Data preprocessing evaluation for web log mining: Reconstruction of activities of a web visitor. *Procedia Computer Science*, *1*(1), 2273–2280. http://dx.doi.org/10.1016/j.procs.2010.04.255

Oztok, M., Zingaro, D., Brett, C., & Hewitt, J. (2013). Exploring asynchronous and synchronous tool use in online courses. *Computers & Education*, *60*(1), 87–94. http://dx.doi.org/10.1016/j.compedu.2012.08.007

Pardos, Z. A., Baker, R. S. J. D., San Pedro, M. O. C. Z., Gowda, S. M., & Gowda, S. M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 117–124. http://dx.doi.org/10.1145/2460296.2460320

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaiane, O. R. (2009). Clustering and sequential pattern mining for online collaborative learning. *IEEE Transactions on Knowledge and Data Engineering*, *21*(6), 759–772. http://dx.doi.org/10.1109/TKDE.2008.138

Raju, G. T., & Satyanarayana, P. S. (2008). Knowledge discovery from web usage data: Complete preprocessing methodology. *International Journal of Computer Science and Network Security*, *8*(1), 179–186.

Roberge, D., Rojas, A., & Baker, R. (2012). Does the length of time off-task matter? *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 234–237. http://dx.doi.org/10.1145/2330601.2330657

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, *21*(1), 135–146. http://dx.doi.org/10.1002/cae.20456

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(6), 601–618. http://dx.doi.org/10.1109/TSMCC.2010.2053532

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368–384. http://dx.doi.org/10.1016/j.compedu.2007.05.016

Romero, M., & Barbera, E. (2011). Quality of learners' time and learning performance beyond quantitative time-on-task. *International Review of Research in Open and Distance Learning*, *12*(5), 125–137. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/999/1870

Rosen, L. D., Mark Carrier, L., & Cheever, N. A. (2013). Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior*, *29*(3), 948–958. http://dx.doi.org/10.1016/j.chb.2012.12.001

Smeets, E., & Mooij, T. (2000). Time on task, interaction, and information handling in multimedia learning environments. *Journal of Educational Computing Research*, *21*(4), 487–502. http://dx.doi.org/10.2190/3KE4-P9E7-L9X8-EMC1

Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational Researcher*, *9*(11), 11–16. http://dx.doi.org/10.3102/0013189X009011011

Swan, K., & Ice, P. (2010). The community of inquiry framework ten years later: Introduction to the special issue. *The Internet and Higher Education*, *13*(1–2), 1–4. http://dx.doi.org/10.1016/j.iheduc.2009.11.003

Trigwell, K., Prosser, M., & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education*, *37*(1), 57–70. http://dx.doi.org/10.1023/A:1003548313194

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology*, *27*(4), 551–572. http://dx.doi.org/10.1016/S0361-476X(02)00006-1

Wise, A. F., Speer, J., Marbouti, F., & Hsiao, Y.-T. (2013). Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviors. *Instructional Science*, *41*(2), 323–343. http://dx.doi.org/10.1007/s11251-012-9230-9

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2013). Learning analytics for online discussions: A pedagogical model for intervention with embedded and extracted analytics. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 48–56. http://dx.doi.org/10.1145/2460296.2460308

Worthen, B. R., Van Dusen, L. M., & Sailor, P. J. (1994). A comparative study of the impact of integrated learning systems on students' time-on-task. *International Journal of Educational Research*, *21*(1), 25–37. http://dx.doi.org/10.1016/0883-0355(94)90021-3

5.3 Summary

All forms of learning take time. With the large body of research suggesting that the amount of time spent on learning can improve the quality of learning as represented by academic performance (Bloom, 1974; Stallings, 1980; Karweit, 1982, 1984), the estimation of students time spent on learning activities is essential. This chapter provides valuable insights into the methodological challenges surrounding estimation of time-on-task measures, and how they overall affect the findings of learning analytics models.

Although within the field of learning analytics, estimation of students time-on-task is relatively common, problems related to time-on-task estimation are rarely described in detail, and the consequences entailed are not fully examined. In this regard, the study presented in this chapter provides some of the first insights into the effects of different time-on-task estimation strategies on the final results of statistical analyses. Our results indicate that the choice of a particular time-on-task estimation strategy plays an important role in the overall model fit and subsequent model interpretations. Hence, in the same manner in which Karweit and Slavin (1982) urged educational researchers of the 1980s to pay attention to the challenges of time-on-task estimation in brick and mortar classrooms, we want to draw the attention of learning analytics researchers of the present day to the similar issues with digital trace data.

In regards to the cognitive presence model described in Chapter two, the present chapter is primarily focused on the model's third, assessment framework layer and the methodological challenges of evidence model operationalization. In particular, the present study concentrates on the development of a valid list of course engagement measures that are related to students' time spent on different learning activities (i.e., evidence rules component of evidence model shown in Figure 3). The study also investigated the sensitivity of a data modeling method to the variability of a key input measure which is critical given the challenges of accurate estimation of student course engagement (i.e., stat model component of evidence model shown in Figure 3). Finally, with the growing importance of study replication on validation of published research findings, the precise specification of all quantitative measures and analytical models is essential. With respect to cognitive presence construct, the assessment model presented in Chapter two offer a potential approach to providing detailed specification of analytic models in a way which is easier to replicate and validate.

6 Validation of the Community of Inquiry model within MOOC context

> It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.
>
> — Arthur Conan Doyle, *A Scandal in Bohemia*

6.1 Introduction

Although the CoI model was originally developed as a conceptual framework focused on inquiry-based learning, the generalizability of the model resulted in its much broader use to model student learning experience in a variety of online and blended learning settings (Arbaugh et al., 2008; Gorsky, Caspi, Blau, Vine, & Billet, 2011). As summarized by Shea and Bidjerano (2010), the CoI model represents the *"concise descriptive model for understanding higher education online learning within an epistemic engagement pedagogical approach."* (p. 1723) As such, the model gained significant traction in the research community, with many considering it one of the best models of online and distance learning to date (Akyol & Garrison, 2011a; Shea & Bidjerano, 2010; Jézégou, 2010).

Given its social-constructivist views on learning and teaching, the CoI model presupposes strong student-teacher interactions and active instructional presence in the course to address particular course objectives, diagnose misconceptions, and to facilitate online discussions in a productive manner (Garrison et al., 1999; Anderson et al., 2001; Garrison & Arbaugh, 2007). Previous studies pointed to the critical role of teaching presence for students' cognitive presence development, and for establishing a social climate that is supportive of open communication and which contributes to the sense of mutual trust and group cohesion (Garrison, Anderson, & Archer, 2010). By examining the relationship between three presences, Garrison, Cleveland-Innes, and Fung (2010) found a direct effect of teaching presence on cognitive presence, the relationship which is also mediated by students' positive social presence.

Because of the need for strong student-teacher interactions, CoI and similar social-constructivist models are rarely used in practice for student cohorts with more than 30-40 students (Anderson

& Dron, 2010). One of the key reasons is that with larger student cohorts, the volume of student interactions is simply too high for instructors to manage, which makes their facilitation much more problematic. This is particularly important for the adoption of the CoI model within novel models of online learning, such as MOOCs, in which cohorts can even have tens or hundreds of thousands of students (OCR, 2015). One potential approach to addressing this problem is the adoption of automated learning analytics systems, which can aid instructors in their instructional activities by providing the assessment of student learning progress, as well as directing their attention to the places where their support is most needed.

While analytics can aid instructors in managing large online courses, it is important to examine the suitability of those models and their pedagogical assumptions in new educational settings. As specifics of each learning context play a significant role in assessing the student learning experience, the first step is to examine the broader applicability of the CoI model for modeling student learning experience within those new contexts. In this thesis, we focus on the use of learning analytics for assessing students' learning within MOOCs as one of the primary models of new models of online learning at scale. As such, the first step is to examine whether the CoI model can be used to support student learning in MOOCs and what key differences between traditional online and MOOC settings with regards to three CoI presences are.

In this chapter, we present the results of a study which examined the use of the CoI model within five MOOC courses using the responses to the CoI survey instrument (Arbaugh et al., 2008). Using the data of some 1,487 students, we examined the hypothesized three-factor structure of the CoI survey instrument, and also the optimal factor structure based on the collected data. Our results indicate that the overall structure of the CoI model is preserved within the MOOC context and that it can be used to adequately capture student learning experience. However, our results also highlight important differences between conventional online and MOOC contexts, in particular in connection to course design, affective expression, and development of cognitive presence.

6.2 Publication: Exploring communities of inquiry in massive open online courses

The following section includes the verbatim copy of the following publication:

> Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., Čukić, I., de Vries, P., Hatala, M., Dawson, S., Siemens, G., and Gašević, D. (2017). Exploring communities of inquiry in massive open online courses. *Manuscript submitted for publication*

Exploring Communities of Inquiry in Massive Open Online Courses

Vitomir Kovanović[a,*], Srećko Joksimović[a], Oleksandra Poquet[b], Thieme Hennis[c], Iva Čukić[a], Pieter de Vries[c], Marek Hatala[d], Shane Dawson[b], George Siemens[e], Dragan Gašević[a]

[a]University of Edinburgh, Old College, South Bridge, Edinburgh EH8 9YL, United Kingdom
[b]University of South Australia, City West campus 217/243 Hindley St, Adelaide, SA 5000, Australia
[c]Delft University of Technology, Jaffalaan 5, 2628 BX, Delft, The Netherlands
[d]Simon Fraser University, Central City Mall, 250-13450 102nd Avenue, Surrey, BC V3T 0A3, Canada
[e]The University of Texas at Arlington, 246 Nedderman Hall, Arlington, TX 76019-0012, United States

Abstract

This study presents an evaluation of the Community of Inquiry (CoI) survey instrument developed by Arbaugh et al. (2008) within the context of Massive Open Online Courses (MOOCs). The study reports the results of a reliability analysis and exploratory factor analysis of the CoI survey instrument using the data of 1,487 students from five MOOC courses. The findings confirmed the reliability and validity of the CoI survey instrument for the assessment of the key dimensions of the CoI model: teaching presence, social presence, and cognitive presence. Although the CoI survey instrument captured the same latent constructs within the MOOC context as in the Garrison's three-factor model (Garrison et al., 1999), analyses suggested a six-factor model with additional three factors as a better fit to the data. These additional factors were 1) course organization and design (a sub-component of teaching presence), 2) group affectivity (a sub-component of social presence), and 3) resolution phase of inquiry learning (a sub-component of cognitive presence). The emergence of these additional factors revealed that the discrepancies between the dynamics of the traditional online courses and MOOCs affect the student perceptions of the three CoI presences. Based on the results of our analysis, we provide an update to the famous CoI model which captures the distinctive characteristics of the CoI model within the MOOC setting. The results of the study and their implications are further discussed.

Keywords: Community of inquiry model, Massive open online courses, Online learning, Exploratory factor analysis

1. Introduction

The growing interest in MOOCs and online education more broadly has been fueled by various social, economic, and political factors that have converged to emphasize the growing societal need for an accessible and sustainable higher education. Some of the factors include concerns surrounding student debt (Matthews, 2013), increasing requirements for lifelong learning to sustain future employment opportunities (Fini, 2009), and an overall need to provide more accessible and democratized models of higher education (Siemens, 2013). While MOOCs have brought online learning to the center of public interest (Kovanović et al., 2015b; Gašević et al., 2014), their development has not been without its challenges.

A particularly important challenge associated with the MOOC development relates to the present state of MOOC pedagogical designs and the disconnect with the current state of research in online and distance education. MOOCs were originally developed by researchers in online education as an experimentation platform for novel online pedagogical approaches based on the constructivist learning theory[1]. A prevalent group of current MOOCs (so-called xMOOCs) have tended to adopt a learning design structured around the pre-recorded video lectures, automated assignments, and quizzes with limited direct teaching interaction undertaken by the instructor. This model of design and teaching is selected for its capacity to scale content and learning activities to a large number of students while diminishing the constraints associated with the need for instructors to engage with individual learners.

The present models of MOOC pedagogical design are essentially focused on the transmission of content. This approach represents a radical departure from contemporary distance education practice that is grounded in social constructivist models of learning (Anderson and Dron, 2010). These models assume that students – rather than assimilating predefined knowledge – actively

*Corresponding author.
Email addresses: v.kovanovic@ed.ac.uk (Vitomir Kovanović), s.joksimovic@ed.ac.uk (Srećko Joksimović), oleksandra.skrypnyk@mymail.unisa.edu.au (Oleksandra Poquet), thieme@hennis.nl (Thieme Hennis), icukic@exseed.ed.ac.uk (Iva Čukić), Pieter.deVries@tudelft.nl (Pieter de Vries), mhatala@sfu.ca (Marek Hatala), Shane.Dawson@unisa.edu.au (Shane Dawson), gsiemens@uta.edu (George Siemens), dragan.gasevic@ed.ac.uk (Dragan Gašević)

[1]This form of MOOCs is now commonly known as connectivist MOOCs or cMOOCs

construct their knowledge through a series of interactions with learning content, instructors, and other students. This knowledge construction process is dependent also on their existing knowledge and experiences, meta-cognitive processes, and a particular learning context. By following the behaviorist notion of learning, the dominating MOOC design arguably represents a step back in the quality and richness of online instruction according to some authors (Stacey, 2013; Bali, 2014). A plausible rationale for this disconnect lies in the multidisciplinary nature of MOOC and online learning research and the strong fragmentation of the MOOC research community to researchers from the field of education and researchers from the field of computer science (Gašević et al., 2014). With researchers from computer science and engineering fields often following a theory-agnostic philosophy of data analysis (Chris et al., 2008), the departure from the contemporary learning theories is not surprising. The disconnect with the previous line of research in online and distance education may also explain the enthusiasm of the early xMOOCs proponents. Although being dubbed a "revolution" (Friedman, 2012) and "tsunami" (Hennessy, 2012) in the field of education, they represent a logical "evolutionary" step in the development of online and distance learning (Bali, 2014; Daniel, 2014).

This paper presents the results of a study examining the use of the contemporary social constructivist models of online and distance education within the MOOC context. The focus of the analysis is on the Community of Inquiry (CoI) model (Garrison et al., 1999), a well-known and one of the most widely-adopted models of distance education (Garrison and Arbaugh, 2007). The CoI model outlines critical dimensions which shape students' online learning experience and also provides a survey instrument used for their assessment (Arbaugh et al., 2008). This paper examines if the CoI survey instrument can be used to evaluate the interactions in MOOC courses. Given the many pedagogical differences between MOOCs and "traditional," small-scale online-courses, a re-validation of the existing CoI survey instrument and its factor structure was conducted using the data of 1,487 students from five MOOC courses. By examining the CoI model of online learning within the MOOC context, we aim to bridge the gap between research in online learning and current MOOC pedagogical practices and to enable its use for assessment of the quality of MOOC learning experience. The results of our analyses and the broader theoretical and practical implications are further discussed.

2. Background work

2.1. Overview of the Community of Inquiry model

The Community of Inquiry (CoI) (Garrison et al., 1999) framework is a widely adopted pedagogical model that outlines the critical dimensions that shape a students' online learning experience. Built upon a social constructivist model of learning, the CoI model (Fig. 1) focuses on the development of higher-order thinking through inquiry-based learning in a learning community. In this context, a learning community is defined as *"a group of individuals who collaboratively engage in purposeful critical discourse and reflection to construct personal meaning and confirm mutual understanding"* (Garrison, 2011, p. 2). The model defines three central dimensions of online learning, known as presences:

1) *Cognitive presence* is a central component in the model which describes phases of inquiry-based learning, including problem conceptualization, knowledge exploration, synthesis, and eventual solution (Garrison et al., 2001).

2) *Social presence* focuses on the important aspects that shape social climate in the course learning community, including student interactivity, group cohesion, and affectivity (Rourke et al., 1999).

3) *Teaching presence* describes different instructional activities before and during the course, which include course organization and design, direct instruction, and facilitation (Anderson et al., 2001).

To evaluate the levels of the three CoI presences, researchers typically employ a self-reported survey instrument (Arbaugh et al., 2008) used to measure the perceived levels of the three presences among the cohort of learners. The CoI model and the CoI survey instrument have been widely used in practice (Garrison et al., 2010a), and have been validated in several research studies (Arbaugh et al., 2008; Gorsky et al., 2011; Rourke and Anderson, 2004). Although originally focused on inquiry-based learning in fully online environments, the generalizability of the CoI model resulted in its wider adoption in the online and blended learning contexts (Garrison et al., 2010a). As such, it has been used as a general framework for assessing students' learning experience within a broad range of learning settings (Anderson and Dron, 2010; Swan and Ice, 2010).

2.2. Community of Inquiry instrument

The CoI survey instrument, originally developed by Arbaugh et al. (2008), consists of thirty-four 5-point Likert scale items designed to measure student perceived levels of teaching (questions 1-13), social (questions 14-22), and cognitive (questions 23-34) presence. As with any survey instrument, the first two issues are whether it is reliable and valid (Field et al., 2012). Reliability concerns whether the instrument provides stable and consistent results (e.g., would similar participants produce similar responses) while validity examines whether the instrument measures what it was designed to measure (Tabachnick and Fidell, 2007). Reliability of the instrument is usually evaluated through Cronbach's α measure (Cronbach, 1951), whereas validity is typically assessed using principal component analysis (PCA) and exploratory factor analysis (EFA) (Field et al., 2012). Both PCA and EFA extract (usually a small) set of latent factors, also called components, which are associated with individual survey questions (Field et al., 2012). If the instrument used to measure N constructs is valid, then PCA or EFA should also reveal N latent factors, which are correctly associated with survey questions (i.e., questions used to measure each construct are all associated with the same factor). With this in mind, several studies examined the reliability and validity of the CoI survey instrument (Arbaugh et al., 2008; Swan et al., 2008; Díaz et al., 2010; Shea and Bidjerano, 2009; Garrison et al., 2010b; Kozan and Richardson, 2014b).

In their seminal study, Arbaugh et al. (2008) conducted a PCA analysis on the data ($N = 287$) from graduate-level courses from four institutions in the USA and Canada. The results of the Arbaugh et al. (2008) study indicated the valid three-factor solution of the CoI survey instrument. An examination of the same dataset by Swan et al. (2008) revealed a strong internal consistency of the CoI survey instrument, with Cronbach's α of .94, .91, and .95 for teaching, social, and cognitive presences, respectively. The PCA analysis was also used by Díaz et al. (2010) for analysis of the data from a both graduate and undergraduate courses at four different institutions ($N = 412$), which provided further confirmation of the CoI instrument reliability and three-factor structure. The only departure from the hypothesized factor structure was related to item #22 (measuring group cohesion in social presence) which loaded almost identically to both social and teaching presence factors (the absolute difference between factor loadings was .004). The reliability and tree factor structure were also confirmed by Shea and Bidjerano (2009), who used EFA on a large dataset ($N = 2,159$) from a multi-institutional fully-online learning program. Similar results using EFA are presented by Garrison et al. (2010b), who analyzed the data from fourteen courses in two study programs ($N = 205$), and by Kozan and Richardson (2014b) who analyzed data ($N = 219$) from students enrolled in a fully online graduate degree program. Similar to the Díaz et al. (2010) study, Kozan and Richardson (2014b) also reported item #22 loading on both the social presence and cognitive presence factors.

It should be noted that studies by Arbaugh et al. (2008) and Díaz et al. (2010) suggested the existence of a potential fourth factor which encompasses survey items related to the course organization and design, a sub-component of the teaching presence. As indicated by Arbaugh et al. (2008), the presence of the fourth factor does not invalidate the theoretical foundations on which the CoI model was developed, as the CoI model theorizes that each of the presences comprises a number of sub-components. For example, teaching presence is defined as consisting of course organization and design, facilitation, and direct instruction, while social presence consists of an affective expression, open communication, and group cohesion (Garrison et al., 1999). The existing literature (Arbaugh, 2007; Shea et al., 2006) also points out to the possibility that teaching presence activities before (i.e., course organization and design) and during the course (i.e., facilitation and direct instruction) might be driven by different dynamics and thus be reflected in the separate factor loadings.

Although the CoI instrument has been used extensively for evaluation of traditional for-credit online and blended learning settings, its adoption in the MOOC context has been limited. To the best of our knowledge, only the study by Damm (2016) used the CoI survey instrument to evaluate the learning experience of students from eight "MOOC-like" non-credit courses offered by a respected U.S. book publisher. However, unlike most MOOCs, these courses had a \$175-\$200 course registration fee, and as a result, were much smaller (around 400 students each). Likewise, it is reasonable to assume that students in these courses had commitments more similar to the traditional for-credit online courses than typical MOOCs which do not charge a registration fee. Although Damm (2016) used the CoI survey instrument to measure the course experience of course participants, they did not evaluate the factor structure of the CoI instrument and instead used in-depth interviews with the students to validate survey results. Overall, their findings suggest that the CoI survey can be used to assess the level of student course engagement and three CoI presences (Damm, 2016).

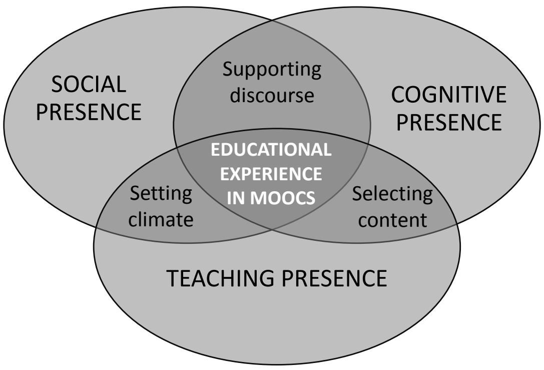


Fig. 1. The original CoI model by Garrison et al. (1999), showing the interconnected nature of the three presences in shaping students' online learning experience.

3. Research questions

While there has been substantial work on the validation of the CoI instrument, the primary context was traditional, formal education, with data coming from the small-scale, for-credit online courses. However, to our knowledge, the use of CoI model and validation of its survey instrument have not been examined within the MOOC context. Given the rapidly emerging MOOC research, as well the broad adoption of the CoI model within traditional online settings, the goal of the present study is to examine whether the CoI survey instrument can be used for examination of student learning experiences within MOOC courses. As the CoI instrument has not been validated in the MOOC context, we focused our investigation on examining its reliability and validity within the MOOC setting. Hence, we focused on the following research questions:

RQ 1: What is the reliability of the Community of Inquiry survey instrument in the MOOC context?

Given that CoI survey instrument has been originally designed for small-scale online learning environments, we first want to examine the reliability of the existing instrument on measuring the levels of the three CoI presences. With many differences between MOOCs and small-scale online courses, it might be that the reliability of the existing instrument is not sufficiently high to reliably measure the three key constructs of the CoI model.

RQ 2: Does the validity of the CoI instrument hold in MOOC setting?

While reliability analyses provide an indication of the internal consistency of the instrument, it is also important to examine the relationship between survey items and the underlying factors which they are supposed to measure. As such, the focus of this question is to examine if the pedagogical differences between traditional online courses and MOOCs impact the validity of the CoI survey instrument and if so, to what extent. For example, it might be that due to specifics of MOOCs, certain questions are not interpreted as originally intended, and thus, reflect different latent constructs than originally theorized.

4. Material and methods

4.1. Study data

The data for this study was collected from five different MOOCs offered by the Delft University of Technology in the Netherlands. All courses were delivered using the edX platform during the Fall 2014 term (Table 1). The courses included a range of learning activities such as recorded video materials, reading materials, short multiple-choice quizzes, homework assignments, and online forum discussions (see Hennis et al., 2016). Before the course commencement, all registered students were invited to complete a voluntary pre-course questionnaire. The questionnaire consisted of items related to a student's reasons for enrolling in the course, anticipated level of course commitment, previous domain knowledge, and their perceived importance of the different course resources and tools. On course completion, all students were again invited to complete a post-course survey which also included the 34 items of the original Community of Inquiry questionnaire by Arbaugh et al. (2008). The basic course descriptive statistics, as well as the number of the CoI survey responses, are shown in Table 1. In total, 2,446 students completed the post-course survey, with a subset of 1,887 students completing the CoI survey questions.

4.2. Data preparation

Before the analysis, following the work of Shea and Bidjerano (2009), we pre-processed the data to remove the data points not suitable for the factor analysis procedure. We removed all incomplete survey responses (7%) and multivariate outliers with Mahalanobis distance larger than 65.25 ($p < .001$) (9%), as done by Shea and Bidjerano (2009). Finally, we removed all cases with standardized Z-scores above 3.29 on any of the 34 CoI survey items (5%). The final dataset consisted of 1,487 cases, which is a 21% reduction of the original dataset.

Table 1
Basic statistics for courses included in the study.

Course	Weeks	Enrolled students	Certified students	Survey responses
Delft Design Approach	10	13,503	136	69
Introduction to Functional Programming	8	38,029	1,968	992
Introduction to Drinking Water Treatment	10	10,543	281	114
Solving Complex Problems	5	32,424	1,396	463
Technology for Bio-based Products	7	9,606	347	249

4.3. Analysis procedure

To evaluate the use of the CoI survey instrument in the MOOC context, we first conducted a scale reliability analysis using Cronbach's alpha and item-rest correlation analysis (Field et al., 2012). In cases where an instrument is used to measure several related constructs, Cronbach (1951) suggested that an analysis should also be conducted for each of the subscales. Thus, we conducted three separate analyses, one for each of the three presences. We also used item-rest correlation to examine whether the reliability of an instrument can be improved by the exclusion of particular survey items (Field et al., 2012).

Post the reliability analysis, an exploratory factor analysis (EFA) was undertaken following the approach by Shea and Bidjerano (2009). Specifically, we conducted a principal axis factoring (PAF) with oblimin rotation and Kaiser normalization to examine the factor structure arising from student completion of the CoI instrument in the MOOC context. The use of oblimin rotation – instead of orthogonal rotation – was warranted based on the interconnected nature of three CoI presences. With 34 manifest variables, the sample of 1,487 cases more than satisfied the popular sample size criteria. For instance, Tinsley and Kass (1979) suggested 5-10 participants per variable (up to 300 participants). Similarly, Tabachnick and Fidell (2007) suggested the inclusion of a sample size of at least 300 cases, while Comrey (1973) defined samples above 1,000 as excellent. The adequacy of our analysis procedure was further evaluated using Bartlett's test of sphericity and Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy.

The prior research has largely demonstrated a stable three-factor structure (Swan et al., 2008; Díaz et al., 2010; Garrison et al., 2010b; Shea and Bidjerano, 2009; Arbaugh et al., 2008; Kozan and Richardson, 2014b) associated with the CoI survey instrument. Hence, for the first validation of the CoI survey instrument, we extracted the three factors using principal axis factoring (PAF). However, we also evaluated the best factor structure that emerged from the collated data set. To select the number of factors for extraction, we evaluated the scree plot (Cattell, 1966) for the inflection point and Kaiser's criterion of eigenvalues larger than one (Kaiser, 1960), as commonly undertaken in EFA/PCA analysis. Given the often ambiguous results by scree plot analysis and Kaiser rule, we complemented our analysis of the number of factors with parallel analysis (Horn, 1965). This approach is considered to be superior to the former two methods (Zwick and Velicer, 1986).

As both PCA (Swan et al., 2008; Díaz et al., 2010; Arbaugh et al., 2008) and EFA (Garrison et al., 2010b; Shea and Bidjerano, 2009; Kozan and Richardson, 2014b) have been used for the study of the CoI survey instrument, we first evaluated the advantages and disadvantages of both methods. Based on our investigation, we eventually decided to use EFA for several reasons. While both PCA and EFA share many similarities and often produce similar results (Velicer, 1974; Velicer et al., 1982; Jensen, 1983), they fundamentally differ in the way in which they model the relationship between latent and manifest variables (Field et al., 2012). PCA considers all variance among the manifest variables to be a shared variance (known as communality) arising from a set of common latent factors (Winter and Dodou, 2016). In contrast, EFA assumes that – aside from the common factors – each manifest variable has a unique latent factor contributing to the unique portion of its variance (the random or unique portion of variance) (Winter and Dodou, 2016). In practice, the PCA procedure derives a lower-rank representation of the manifest variable covariance matrix, while EFA provides a more sound modelling of the relationship between a set of variables (Field et al., 2012). As such, it is often considered to be the only procedure that can be used to estimate the underlying structure of latent factors (Field et al., 2012). In cases where a unique portion of variance is small and manifest variables strongly load on a single factor, both methods produce similar results (Winter and Dodou, 2016; Guadagnoli and Velicer, 1988). However, in cases with smaller communality, the differences can be more significant as pointed out by Snook and Gorsuch (1989), and Widaman (1990, 1993).

Furthermore, while results of both PCA and EFA depend on the number of extracted factors, they differ in ways in which they respond to over- and under-extraction of latent factors (Winter and Dodou, 2016). Although in both methods under-extraction is a more serious problem than over extraction (Fava and Velicer, 1996), it is shown that results of PCA are more severely distorted due to the over-extraction (Lawrence and Hancock, 1999). Given the unexplored nature of the CoI survey instrument in the MOOC context, this is another reason to favor EFA over PCA.

5. Results

5.1. RQ1: Reliability analysis results

To validate the CoI survey instrument in the MOOC context, we examined the reliability of the CoI instrument using Cronbach's alpha measure (Cronbach, 1951). All three subscales obtained overall reliability scores of .89 or above (Table 2) which indicates a reliable measurement instrument (Kline, 1999). We can also see that none of the items on all three subscales had an alpha value higher than the overall alpha value, indicating that none of the items negatively affects instrument reliability. Likewise, the correlations of all item with the rest of their respective scale items were sufficiently high, i.e., significantly above the threshold of .3 used in the literature (Field et al., 2012; Everitt, 2002).

5.2. RQ2: Exploratory factor analysis results

An exploratory factor analysis (EFA) using the principal axis factoring was conducted on the 34 items of the community of inquiry survey instrument collected from 1,487 MOOC participants. The Kaiser-Meyer-Olkin (KMO) measure confirmed the

Table 2
Reliability analysis results.

Teaching presence			Social presence			Cognitive presence		
Item	Cronbach's α	Item-rest r	Item	Cronbach's α	Item-rest r	Item	Cronbach's α	Item-rest r
Overall	.93	.72	Overall	.89	.67	Overall	.9	.65
TP1	.94	.69	SP1	.89	.58	CP1	.9	.66
TP2	.94	.71	SP2	.9	.52	CP2	.9	.68
TP3	.93	.71	SP3	.9	.54	CP3	.9	.72
TP4	.94	.59	SP4	.88	.71	CP4	.91	.61
TP5	.93	.76	SP5	.88	.77	CP5	.91	.63
TP6	.93	.77	SP6	.87	.81	CP6	.91	.47
TP7	.93	.77	SP7	.88	.72	CP7	.9	.7
TP8	.93	.8	SP8	.89	.66	CP8	.9	.73
TP9	.94	.6	SP9	.88	.7	CP9	.9	.7
TP10	.93	.75				CP10	.9	.68
TP11	.93	.78				CP11	.91	.61
TP12	.94	.69				CP12	.91	.62
TP13	.94	.68						

adequacy of our sample, KMO=0.95 ("superb" according to Kaiser (1974)), with all individual KMO scores above 0.86 (Table 3) which is higher than the accepted threshold of 0.5 (Kaiser, 1974). The results of Bartlett's test were highly significant $\chi^2(561) = 34,045.36$, $p < .00001$. The results of these analyses, together with the satisfaction of the popular sample size criteria (Tinsley and Kass, 1979; Tabachnick and Fidell, 2007; Comrey, 1973), provide sufficient validation of the adequacy of our sample.

To select the optimal number of factors for extraction, we first plotted the eigenvalues from the PCA analysis (Fig. 2 and Table 4). The scree plot gave inconclusive results with either five or six factors being an optimal solution. However, both Kaiser rule (Kaiser, 1960) of eigenvalues larger than one and parallel analysis (Horn, 1965) indicated an optimal the six-factor model. In the rest of the analysis, we focused on the original three-factor model and the discovered six-factor model.

5.2.1. Three factor model results

To confirm the original structure of the CoI survey instrument, we first examined the three-factor solution of the principal axis factoring with oblimin rotation and Kaiser normalization. Overall, the model accounted for 52% of the variance (Table 5), with the average item communality of 0.52. The fit based on the off-diagonal values was 0.99, and the root mean square of the residuals (RMSR) was 0.05. These results are indicators of an overall good model fit (Field et al., 2012).

Table 3
Individual KMO scores. Overall KMO score: .95.

Teaching presence		Social presence		Cognitive presence	
Question	KMO	Question	KMO	Question	KMO
TP1	.94	SP1	.86	CP1	.95
TP2	.94	SP2	.86	CP2	.93
TP3	.96	SP3	.94	CP3	.94
TP4	.96	SP4	.94	CP4	.92
TP5	.97	SP5	.88	CP5	.96
TP6	.96	SP6	.89	CP6	.97
TP7	.96	SP7	.93	CP7	.95
TP8	.97	SP8	.94	CP8	.96
TP9	.97	SP9	.96	CP9	.97
TP10	.97			CP10	.95
TP11	.97			CP11	.95
TP12	.94			CP12	.94
TP13	.94				

Table 4
Eigenvalues from principal component analysis.

	Initial eigenvalues		
Component	Total	Percentage of variance	Cumulative percentage
1	12.62	.37	.37
2	3.99	.12	.49
3	2.56	.07	.56
4	1.39	.04	.60
5	1.12	.03	.64
6	1.09	.03	.67

The three factors explained 21%, 17%, and 14% of the variability, respectively (Table 5). All survey items loaded with 0.3 or more at only one factor, and only three items (item #2, #9, and #28) loaded on two factors with 0.2 or more (Table 5). The item clustering suggests that the first factor represents teaching presence (`TP`), the second factor cognitive presence (`CP`), and the third factor social presence (`SP`). Correlation analyses among the factors given in Table 6a revealed a strong correlation between cognitive and teaching presences (0.61), while social presence correlated moderately with both teaching presence (0.33) and cognitive presence (0.34). The individual item loadings unveiled that all but one item (question #28) loaded significantly only on their hypothesized factors. Question #28, related to the exploration phase of cognitive presence, had a standardized loading of .37 with the factor representing social presence, and .26 to the factor representing cognitive presence.

Following the initial three-factor analysis of the CoI survey instrument, we conducted an additional analysis without item #28 which was shown to load on a factor representing social presence (Table 5). Although loadings in an EFA analysis are more resistant to additions and removals of survey items than PCA (Widaman, 2007), we wanted to confirm whether there were any significant changes in the factor structure. The changes in item loadings and the overall model statistics were minor (on the second decimal point), indicating that the inclusion of the particular survey item did not negatively affect the results of our factor analysis.

5.2.2. Six factor model results

In addition to the original three-factor model, we also examined the six-factor model which was shown to provide the best fit for our data. The six-factor model accounted for 61% of the variance, 9% more than the original three-factor solution (Table 5) with the average item communality of .61. The fit based on the off-diagonal values was higher than .99, and the root mean square of the residuals (RMSR) was .02, indicating an excellent model fit (Field et al., 2012).

The six factors accounted for 16%, 13%, 12%, 7%, 7%, and 5% of the variance, respectively (Table 5). The grouping of survey items suggests that the first factor related primarily to teaching presence (`TP`), except the course organization and design which was captured by the fifth factor (`Org.`). The second factor primarily related to the cognitive presence (`CP`), except for the resolution phase which was captured by the sixth factor (`Res.`). Finally, the third factor related to the social presence (`SP`), except for the first two social presence items which were related to the level of affective expression between students (`Aff.`). The correlations between the extracted factors are shown in Table 6b. As expected, the three new factors most significantly correlated with the factor

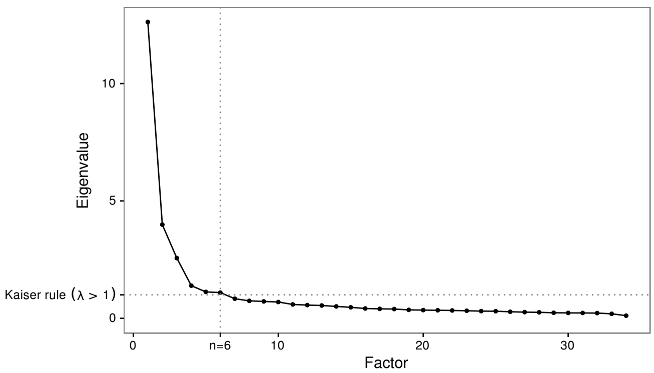


Fig. 2. Principal component analysis scree plot.

Table 5
Factor loading matrix for the three- and six-factor models. The biggest loading for each item is shown boldface.

		Three factor model			Modified three factor model			Six-factor model					
#	Question	TP	CP	SP	TP	CP	SP	TP	CP	SP	Res.	Org.	Aff.
1.	The teaching team clearly communicated important course topics.	**.63**	.15	-.12	**.63**	.15	-.11	.07	-.03	-.01	.06	**.64**	.03
2.	The teaching team clearly communicated important course goals.	**.62**	.17	-.10	**.62**	.17	-.09	.01	.002	.02	.02	**.70**	.03
3.	The teaching team provided clear instructions on how to participate in course learning activities.	**.67**	.08	-.06	**.66**	.08	-.06	.17	.002	.01	-.02	**.56**	.04
4.	The teaching team clearly communicated important due dates/time frames for learning activities.	**.53**	.10	-.02	**.52**	.11	-.02	.20	.08	.03	-.04	**.36**	-.01
5.	The teaching team was helpful in identifying areas of agreement and disagreement in course discussions.	**.80**	-.09	.02	**.80**	-.08	.02	**.58**	-.01	-.001	-.03	.22	.01
6.	The teaching team was helpful in guiding the class towards understanding course topics	**.77**	.02	-.05	**.76**	.03	-.05	**.42**	-.03	.02	.03	.39	-.04
7.	The teaching team helped to keep course participants engaged and participating in productive dialogue.	**.82**	-.09	.02	**.82**	-.09	.02	**.71**	.02	.01	-.04	.11	-.06
8.	The teaching team helped keep the course participants on task in a way that helped me to learn.	**.85**	-.08	.01	**.85**	-.08	.004	**.66**	.05	-.03	-.08	.20	-.005
9.	The teaching team encouraged course participants to explore new concepts in this course.	**.50**	.20	-.10	**.50**	.20	-.09	**.38**	.26	-.01	-.08	.15	-.16
10.	Instructor teaching team reinforced the development of a sense of community among course participants.	**.79**	-.13	.10	**.80**	-.12	.10	**.80**	-.09	-1e-04	.11	-.01	.02
11.	The teaching team helped to focus discussion on relevant issues in a way that helped me to learn.	**.77**	-.03	.07	**.78**	-.03	.07	**.73**	-.01	.01	.10	.04	-.01
12.	The teaching team provided feedback that helped me understand my strengths and weaknesses.	**.75**	-.15	.12	**.75**	-.14	.12	**.77**	-.05	-.06	.05	-.05	.11
13.	The teaching team provided feedback in a timely fashion.	**.71**	-.13	.13	**.71**	-.13	.13	**.78**	.02	.03	-.04	-.08	-.01
14.	Getting to know other course participants gave me a sense of belonging in the course.	.15	-.10	**.55**	.16	-.08	**.54**	-.03	-.04	-.04	.01	.08	**.91**
15.	I was able to form distinct impressions of some course participants.	.09	-.07	**.51**	.10	-.06	**.50**	.001	.03	-.01	-.02	.01	**.74**
16.	Online or web-based communication is an excellent medium for social interaction.	.06	.06	**.50**	.06	.07	**.49**	-.03	.05	**.34**	-.01	.10	.22
17.	I felt comfortable conversing through the online medium.	-.10	.09	**.73**	-.10	.11	**.73**	-.10	.03	**.81**	-.03	.08	-.10
18.	I felt comfortable participating in the course discussions.	-.11	.04	**.84**	-.11	.07	**.84**	-.08	-.05	**.94**	.02	.06	-.13
19.	I felt comfortable interacting with other course participants.	-.11	.02	**.88**	-.11	.05	**.89**	-.08	-.06	**.95**	.01	.05	-.09
20.	I felt comfortable disagreeing with other course participants while still maintaining a sense of trust.	-.08	-.001	**.80**	-.08	.02	**.80**	.07	-.03	**.81**	.02	-.10	-.09
21.	I felt that my point of view was acknowledged by other course participants.	-.02	-.02	**.71**	-.01	-.002	**.71**	.06	-.03	**.60**	.02	-.05	.09
22.	Online discussions help me to develop a sense of collaboration.	-.01	2e-04	**.72**	.003	.02	**.72**	.07	.03	**.54**	.002	-.06	.17
23.	Problems posed increased my interest in course issues.	.03	**.68**	-.04	.03	**.68**	-.03	-.17	**.56**	-.002	.03	.24	.01
24.	Course activities piqued my curiosity.	.001	**.75**	-.08	-.001	**.75**	-.07	-.23	**.63**	-.05	.01	.27	.02
25.	I felt motivated to explore content related questions.	-.07	**.82**	-.03	-.07	**.82**	-.02	-.18	**.69**	-.01	.03	.15	-.01
26.	I utilized a variety of information sources to explore problems posed in this course.	-.12	**.69**	.01	-.11	**.69**	.02	-.01	**.75**	.01	-.08	-.11	-.08
27.	Brainstorming and finding relevant information helped me resolve content related questions.	.02	**.57**	.07	.03	**.56**	.07	.15	**.63**	-.08	-.004	-.17	.09
28.	Online discussions were valuable in helping me appreciate different perspectives.	.08	.25	**.37**	-	-	-	.23	**.40**	.19	-.10	-.17	.11
29.	Combining new information helped me answer questions raised in course activities.	-.03	**.67**	.08	-.01	**.66**	.07	.08	**.71**	-.02	-.02	-.12	.03
30.	Learning activities helped me construct explanations/solutions.	.05	**.70**	.03	.06	**.69**	.03	.03	**.59**	.01	.09	.04	-.003
31.	Reflection on course content and discussions helped me understand fundamental concepts in this class.	.03	**.66**	.05	.03	**.66**	.05	.08	**.49**	.05	.19	-.03	-.05
32.	I can describe ways to test and apply the knowledge created in this course.	.004	**.69**	-.005	.002	**.69**	.005	.01	.10	.04	**.75**	-.004	-.05
33.	I have developed solutions to course problems that can be applied in practice.	.04	**.58**	.02	.04	**.58**	.03	.10	.01	-.01	**.77**	-.07	.02
34.	I can apply the knowledge created in this course to my work or other non-class related activities.	.03	**.62**	-.02	.02	**.63**	-.01	-.05	-.01	-.002	**.80**	.08	.03
	Eigenvalue	7.17	5.79	4.88	7.12	5.67	4.69	5.47	4.28	4.17	2.5	2.34	1.83
	Percentage of variance	.21	.17	.14	.22	.17	.14	.16	.13	.12	.07	.07	.05
	Total variance	.21	.38	.52	.22	.39	.53	.16	.41	.28	.48	.55	.61
	Alpha	.93	.88	.9	.93	.89	.9	.92	.88	.88	.78	.82	.6
	Correlation	.72	.65	.67	.72	.67	.65	.74	.66	.72	.71	.71	.7

Table 6
Correlation between factors in two examined models.

(a)
Correlation between factors for the three-factor solution

	TP	CP	SP
TP	1.00	0.61	0.33
CP		1.00	0.34
SP			1.00

(b)
Correlation between factors for the six-factor solution

	TP	CP	SP	Res.	Org.	Aff.
TP	1.00	0.47	0.29	0.45	0.46	0.34
CP		1.00	0.32	0.64	0.34	0.23
SP			1.00	0.29	0.07	0.50
Res.				1.00	0.35	0.21
Org.					1.00	0.02
Aff.						1.00

representing the rest of their respective subscales (i.e., course organization and design with the teaching presence, group affectivity with the social presence, and resolution phase with the cognitive presence).

6. Discussion

6.1. RQ1: Reliability of the CoI instrument in the MOOC context

The results from the reliability analysis confirmed that the use of the CoI survey instrument within the MOOC context is internally consistent. The obtained Cronbach's α values for the three subscales were just slightly lower than the ones in the existing research (Swan et al., 2008) and still sufficiently above the .8 level which is often used in the literature (Kline, 1999). Similar to the previous studies (Swan et al., 2008; Díaz et al., 2010; Garrison et al., 2010b; Shea and Bidjerano, 2009), the lowest level of internal consistency was achieved for social presence (.89) and the highest for the teaching presence (.93).

6.2. RQ2: Validation of the CoI factor structure in the MOOC context

6.2.1. Validation of the original three-factor model

To identify whether the factor structure of the CoI survey instrument was influenced by the MOOC pedagogical design and learning context, we examined the original three-factor structure and compared it with the existing literature (Swan et al., 2008; Díaz et al., 2010; Garrison et al., 2010b; Shea and Bidjerano, 2009; Arbaugh et al., 2008; Kozan and Richardson, 2014b). The results of the previously published studies and the present study are summarized in Table 7. Our results indicate that the factor structure of the CoI survey instrument still holds in the MOOC context and is aligned with the existing literature (Garrison et al., 2010b; Kozan and Richardson, 2014b). Similarly, our results indicate a strong correlation between teaching and cognitive presences, and a moderate correlation of social presence with both teaching and cognitive presences (Table 6a), which is also aligned with the published studies (Arbaugh et al., 2008; Shea and Bidjerano, 2009; Kozan and Richardson, 2014a)

Overall, the three-factor model explained 52% of variance which is very similar to the results reported by Garrison et al. (2010b) (54%), and slightly lower than reported in other studies in non-MOOC context, which typically reported between 60-65% of the explained variance (Table 6a). However, the percentage of variance explained is highly dependent on the adopted analysis procedure and particular study details. In our case, the results of PCA analysis (Table 4), which we conducted for the purpose of scree-plot analysis, show higher percentages of variance explained than for the EFA analysis (Table 5). The solution with three principal components accounted for 56% of the variance, while the solution with six principal components accounted for 67% of the variance (Table 4). The higher percentage of variance explained for PCA is because it focuses on maximizing variance explained by each subsequent factor (the latent model relations are essentially a side-product), whereas EFA directly models relationships between latent and manifest variables. Similarly, using six latent variables, which was suggested as optimal, could explain more variability in the collected data, than in the cases when only three factors were used.

This study found that one survey item (i.e., question #28 measuring cognitive presence: *"Online discussions were valuable in helping me appreciate different perspectives"*) loaded on the "wrong" factor, i.e., loaded onto social presence. The differences in factor loadings between traditional for-credit online courses and MOOCs suggest a specific relationship between social and cognitive presence within MOOC contexts. These differences in factor loadings are likely a result of the differences in course designs between MOOCs and small-scale online courses which put more emphasis on discussion participation.

In addition to pedagogical differences, there are also substantial differences in the basic demographics of students enrolling in MOOCs and traditional, small scale, for-credit online courses (Hennis et al., 2016). These differences can have a strong influence on the use of the available technologies and tools such as online discussions (Kovanović et al., 2015a). As such, students in MOOCs are likely perceiving discussion participation as more social, rather than a cognitive activity which is likely reflected on the loading of the abovementioned (cognitive presence) item (#28) to be more related to the social than the cognitive presence factor.

6.2.2. Examination of the optimal six-factor model

Interestingly, the model with six factors was suggested as optimal by Kaiser's rule, parallel analysis, and also partially by the scree plot analysis (Fig. 2). The model with six factors explained additional 9% of the variance and distinguished between different sub-components of the three presences. The three added factors were related to the resolution phase of cognitive presence, course organization and design within teaching presence, and affective communication within social presence (Table 5). The differences in the factor structure are likely resulting from the significant pedagogical differences between traditional online courses and MOOCs and emphasize the unique characteristics of MOOCs regarding the development of the communities of inquiry.

Building upon the original diagram of the CoI model by Garrison et al. (1999), we developed an updated visual representation of the CoI model which emphasizes the specifics of communities of inquiry within the MOOC context (Fig. 3). The three smaller inner circles are used to represent the three additional latent factors, and to emphasize the unique characteristics that these three sub-components have within the MOOC setting. While we initially indented to preserve the original Venn diagram notation of the CoI model (Fig. 1), given the precise semantics of Venn diagrams (in particular the meaning of circle overlapping), we decide for the model shown in Fig. 3. However, like in the previous work on the validation of the CoI survey instrument, it is important to point out that the identification of the six-factor model does not invalidate the theoretical foundation of the CoI model or the usability of the CoI survey instrument. While certain survey items (e..g, Item #28) might require some changing and rephrasing in the context of MOOCs, the overall results indicate that the current CoI survey instrument can be used in the MOOC context without raising issues of the instrument's internal consistency or validity.

The fourth factor found by our analysis was associated with the items assessing the levels of resolution within students' cognitive presence development. Loading of these items on a separate factor is an indication that questions related to the first three phases of cognitive presence and the items focused on the final phase of resolution capture two different learning processes. Several reasons are likely contributing to this. As we already know from the literature, students in traditional online courses often fail to reach higher levels of cognitive presence (i.e., integration and resolution) (Garrison et al., 2001). This failure is usually attributed (to a large extent) to the course design and expectations (Garrison et al., 2010a; Gašević et al., 2015). Secondly, the literature also showed the critical importance of teachers' role in reaching the resolution phase (Celentin, 2007; Garrison and Arbaugh, 2007). Finally, a significant impact of time and course duration on the development of three presences has also been suggested (Akyol and Garrison, 2008; Akyol et al., 2011). For instance, Akyol et al. (2011) showed that students in a shorter, six-week version of a course did not reach the integration and resolution to the same extent as the students in a longer, thirteen-week version of the same course. With this in mind, the open nature of MOOCs, their broad accessibility to a diverse student population, the limited direct instruction and facilitation, and the shorter course length than formal for-credit online courses all contribute to the resolution having different dynamics separate from other phases of cognitive presence. While future research is needed on understanding what the driving force behind reaching the resolution within MOOC context is, one possible explanation might be the different motivations for participation in MOOCs, in general, and for participation in online discussions, in particular. For example, it might be that reaching high levels of cognitive presence requires significantly more active forum participation which is not mandated by the course design. Hence, students who engage in active forum participation might be more likely to reach the higher levels of cognitive presence than students focused on the individual learning activities.

The items related to course organization and design loaded on a separate (fifth) factor while the rest of teaching presence items (i.e., facilitation and direct instruction) loaded on the first factor. Loading of teaching presence items onto two different factors is an indication that course organization and design represent a unique construct within MOOC contexts. Similar findings were already reported by Arbaugh et al. (2008) who noted the existence of two factors related to teaching presence, the first one representing

Table 7
Comparison of the present study findings with the existing studies of the CoI survey instrument.

Study	Method	Var.	Factor (var.)	Factor (var.)	Factor (var.)	Factor (var.)	Factor (var.)	Factor (var.)
Arbaugh et al. (2008) and	PCA	61.3%	TP (51%)	SP (5.6%)	CP (4.5%)			
Swan et al. (2008)		64.7%	TP' (51%)	SP (5.6%)	CP (4.5%)	TP" (3.5%)		
Garrison et al. (2010b)	EFA	53.6%	TP (38.5%)	CP (9%)	SP (6.1%)			
Kozan and Richardson (2014b)	EFA	64.8%	TP (48.2%)	CP (10.6%)	SP (6%)			
Shea and Bidjerano (2009)	EFA	64.2%	CP (50.6%)	TP (9.6%)	SP (3.9%)			
Díaz et al. (2010)	PCA	61.9%	CP (44.2%)	TP (10.6%)	SP (7.2%)			
		66.2%	CP (44.2%)	TP' (10.6%)	TP" (7.2%)	SP (4.3%)		
Present study	EFA	52%	TP (21%)	CP (17%)	SP (14%)			
		61%	TP' (16%)	CP' (13%)	SP' (12%)	Res. (7%)	Org. (7%)	Aff. (5%)

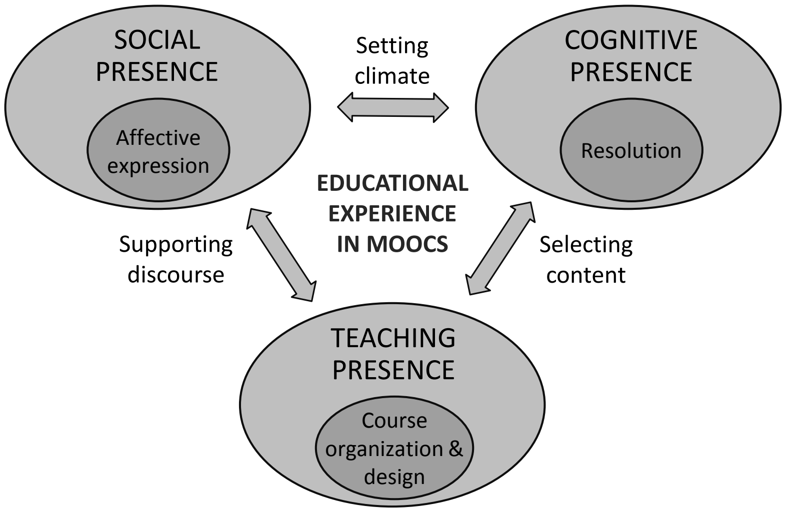


Fig. 3. Updated version of the CoI model by Garrison et al. (1999) which captures the distinct characteristics of the CoI model within the MOOC context. The smaller inner circles emphasize the specifics of course organization & design, affective expression, and resolution phase within the MOOC setting.

course organization and design, and the second one representing facilitation and direct instruction. As suggested by Arbaugh et al. (2008), the separate factors capture the different times at which these teaching activities take place (i.e., organization and design pre-course, and facilitation and direct instruction during the course). In the MOOC context, the difference between teaching activities that happen before and during the course is even more emphasized, as most MOOCs follow a very structured and predefined course organization with almost no changes during the course. Given a massive number of students in a MOOC, even slightest changes are very challenging to implement during the course execution (Jaschik, 2013). Similarly, given the limited teaching staff involved, a majority of MOOCs employ pre-recorded videos for setting up course goals and objectives, and expectations of student course participation. As well, MOOCs frequently use automated methods for feedback and assessment (e.g., computer-graded quizzes and assignments). Although further research is necessary, it is likely that due to this "dehumanization" of the role of the teacher before and during the course manifests as two separate constructs.

Finally, the first two items related to the affectivity group of the social presence loaded on a separate (sixth) factor. This indicates different dynamics surrounding the development of affect in the group communication among the students in a MOOC. In this regard, previous research (Garrison, 2011; Poquet et al., 2016; Akyol and Garrison, 2008) noted a critical importance of time and cohort size on the development of social presence. For example, Akyol et al. (2011) showed that students in a shorter version of a course had significantly lower levels of affective expression than students in the longer version of the same course. Similarly, Poquet et al. (2016) also reported challenges of establishing affective expression in MOOCs, particularly in shorter courses with large student cohorts. Based on this, it seems likely that some of the unique pedagogical characteristics of MOOCs, namely larger student cohorts and shorter course duration, have a significant effect on the development of affectivity in MOOCs as a process separate from social presence.

6.3. Open Questions and Future Work

While the current study provided insights into the use of the CoI instrument in the MOOC context, there are some open questions. In particular, while our study validated the use of the CoI survey instrument, it is also important to understand the effects of and relationships between the three CoI presences. Hence, in our future work we will also investigate the relationships between the three presences, similarly to the work of Shea and Bidjerano (2009) and Garrison et al. (2010b). Given the specifics of learning in MOOCs, it is important to examine whether the existing relationships between the three presences still hold in the MOOC context and whether there are some particular differences in sustaining communities of inquiry within MOOC courses.

7. Conclusions

In this paper, we evaluated the use of the CoI survey instrument within the MOOC context. Through the exploratory factor analysis of the data ($N = 1,487$) from five MOOCs, we examined whether the differences between traditional small-scale online courses, for which the CoI survey was initially designed, and MOOCs affect the reliability and validity of the CoI survey instrument. First of all, our results indicate that *Community of Inquiry survey instrument is a reliable and valid tool for measuring the perceived levels of teaching, cognitive, and social presences within the context of Massive Open Online Courses*. The demonstrated validity and reliability of the CoI instrument are important from the practical perspective as the present instrument can be easily included in the default post-course evaluation surveys which are administered in many MOOCs today. The inclusion of the CoI survey instrument would then enable the examination of how the particular characteristics of the course (e..g, organization and design, subject domain, or student population) affect the levels of three presences. Through the analysis of the relationships between the three presences (Shea and Bidjerano, 2009), the CoI survey data can also provide an improved understanding of the MOOC learning processes. Most importantly, it would enable a pedagogically-sound evaluation and quality assurance of MOOCs.

In addition to the validation of the CoI instrument, the current study also revealed some specifics of the MOOC context which are summarized in the updated CoI model shown on Fig. 3. While our results validated the structure of the three-factor model, all model selection criteria (i.e., scree plot, Kaiser criterion, and parallel analysis) indicated a six-factor model as optimal. The three additional factors correspond to 1) course organization and design (sub-component of teaching presence), 2) resolution phase (sub-component of cognitive presence), and 3) affective expression (sub-component of social presence). These differences highlight the key areas in which the MOOC context is different from the traditional small-scale online course context. Firstly, the open nature of MOOCs, their shorter duration and limited instructor involvement negatively impact reaching the higher levels of cognitive presence. Secondly, given the large number of students and the limited interactions between students and instructors, course organization and design are of particular importance, and represent a construct separated from the rest of teaching presence. Finally, affective expression in student group communication seems especially challenging to develop, which is likely caused by the large student cohorts and shorter course duration (Garrison, 2011; Akyol and Garrison, 2008; Akyol et al., 2011).

References

Akyol, Z., Garrison, D.R., 2008. The Development of a Community of Inquiry over Time in an Online Course: Understanding the Progression and Integration of Social, Cognitive and Teaching Presence. Journal of Asynchronous Learning Networks 12, 3–22. URL: http://eric.ed.gov/?id=EJ837483.

Akyol, Z., Vaughan, N., Garrison, D.R., 2011. The impact of course duration on the development of a community of inquiry. Interactive Learning Environments 19, 231–246. doi:10.1080/10494820902809147.

Anderson, T., Dron, J., 2010. Three generations of distance education pedagogy. The International Review of Research in Open and Distance Learning 12, 80–97. URL: http://www.irrodl.org/index.php/irrodl/article/view/890.

Anderson, T., Rourke, L., Garrison, D.R., Archer, W., 2001. Assessing Teaching Presence in a Computer Conferencing Context. Journal of Asynchronous Learning Networks 5, 1–17. URL: http://onlinelearningconsortium.org/wp-content/plugins/olc-download/download.php?post_id=4190.

Arbaugh, J., Cleveland-Innes, M., Diaz, S.R., Garrison, D.R., Ice, P., Richardson, J.C., Swan, K.P., 2008. Developing a community of inquiry instrument: Testing a measure of the Community of Inquiry framework using a multi-institutional sample. The Internet and Higher Education 11, 133–136. doi:10.1016/j.iheduc.2008.06.003.

Arbaugh, J.B., 2007. An Empirical Verification of the Community of Inquiry Framework. Journal of Asynchronous Learning Networks 11, 73–85. URL: http://eric.ed.gov/?id=EJ842689.

Bali, M., 2014. MOOC Pedagogy: Gleaning Good Practice from Existing MOOCs. Journal of Online Learning and Teaching 10, 44–56. URL: http://jolt.merlot.org/vol10no1/bali_0314.pdf.

Cattell, R.B., 1966. The Scree Test For The Number Of Factors. Multivariate Behavioral Research 1, 245–276. doi:10.1207/s15327906mbr0102_10.

Celentin, P., 2007. Online training: Analysis of interaction and knowledge building patterns among foreign language teachers. Journal of Distance Education 3, 39–58. URL: http://www.ijede.ca/index.php/jde/article/view/29.

Chris, A., Chris, e., Science, e., 2008. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. URL: https://www.wired.com/2008/06/pb-theory/.

Comrey, A.L., 1973. A First Course in Factor Analysis. Academic Press, New York, NY.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334. doi:10.1007/BF02310555.

Damm, C.A.V., 2016. Applying a Community of Inquiry Instrument to Measure Student Engagement in Large Online Courses. Current Issues in Emerging eLearning 3, 138–172. URL: http://scholarworks.umb.edu/ciee/vol3/iss1/9/.

Daniel, J., 2014. Foreword to the Special Section on Massive Open Online Courses MOOCs – Evolution or Revolution? Journal of Online Learning and Teaching 10, i–iv. URL: http://jolt.merlot.org/vol10no1/daniel_foreword_0314.pdf.

Díaz, S.R., Swan, K., Ice, P., Kupczynski, L., 2010. Student ratings of the importance of survey items, multiplicative factor analysis, and the validity of the community of inquiry survey. The Internet and Higher Education 13, 22–30. doi:10.1016/j.iheduc.2009.11.004.

Everitt, B., 2002. Cambridge dictionary of statistics. Second ed., Cambridge University Press.

Fava, J.L., Velicer, W.F., 1996. The effects of underextraction in factor and component analyses. Educational and Psychological Measurement 56, 907–929. doi:10.1177/0013164496056006001.

Field, A.P., Miles, J., Field, Z., 2012. Discovering statistics using R. SAGE, Los Angeles; London.

Fini, A., 2009. The Technological Dimension of a Massive Open Online Course: The Case of the CCK08 Course Tools. The International Review of Research in Open and Distance Learning 10. URL: http://www.irrodl.org/index.php/irrodl/article/view/643.

Friedman, T.L., 2012. Come the Revolution. The New York Times URL: http://www.nytimes.com/2012/05/16/opinion/friedman-come-the-revolution.html.

Garrison, D.R., 2011. E-Learning in the 21st Century: A Framework for Research and Practice. Second ed., Routledge, New York.

Garrison, D.R., Anderson, T., Archer, W., 1999. Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. The Internet and Higher Education 2, 87–105. doi:10.1016/S1096-7516(00)00016-6.

Garrison, D.R., Anderson, T., Archer, W., 2001. Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education. American Journal of Distance Education 15, 7–23. doi:10.1080/08923640109527071.

Garrison, D.R., Anderson, T., Archer, W., 2010a. The first decade of the community of inquiry framework: A retrospective. The Internet and Higher Education 13, 5–9. doi:10.1016/j.iheduc.2009.10.003.

Garrison, D.R., Arbaugh, J., 2007. Researching the community of inquiry framework: Review, issues, and future directions. The Internet and Higher Education 10, 157–172. doi:10.1016/j.iheduc.2007.04.001.

Garrison, D.R., Cleveland-Innes, M., Fung, T.S., 2010b. Exploring causal relationships among teaching, cognitive and social presence: Student perceptions of the community of inquiry framework. The Internet and Higher Education 13, 31–36. doi:10.1016/j.iheduc.2009.10.002.

Gašević, D., Adesope, O., Joksimović, S., Kovanović, V., 2015. Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. The Internet and Higher Education 24, 53–65. doi:10.1016/j.iheduc.2014.09.006.

Gašević, D., Kovanović, V., Joksimović, S., Siemens, G., 2014. Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative. The International Review of Research in Open and Distributed Learning 15. URL: http://www.irrodl.org/index.php/irrodl/article/view/1954.

Gorsky, P., Caspi, A., Blau, I., Vine, Y., Billet, A., 2011. Toward a CoI population parameter: The impact of unit (sentence vs. message) on the results of quantitative content analysis. The International Review of Research in Open and Distance Learning 13, 17–37. URL: http://www.irrodl.org/index.php/irrodl/article/view/1073.

Guadagnoli, E., Velicer, W.F., 1988. Relation of sample size to the stability of component patterns. Psychological Bulletin 103, 265–275. doi:10.1037/0033-2909.103.2.265.

Hennessy, J.L., 2012. The Coming Tsunami in Educational Technology. Keynote presentation at the CRA's 40th Anniversary Conference, Snowbird, UT. URL: http://archive2.cra.org/uploads/documents/resources/snowbird2012_slides/hennesy.pdf.

Hennis, T., Topolovec, S., Poquet, O., de Vries, P., 2016. Who is the Learner: Profiling the Engineering MOOC Student, in: SEFI 44th Annual Conference, Tampere, Finland. URL: http://www.sefi.be/conference-2016/papers/Open_and_Online_Engineering_Education__Engineering_Education_Research/hennis-who-is-the-learner-108_a.pdf.

Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30, 179–185. doi:10.1007/BF02289447.

Jaschik, S., 2013. MOOC Mess. Inside Higher Ed URL: https://www.insidehighered.com/news/2013/02/04/coursera-forced-call-mooc-amid-complaints-about-course.

Jensen, A.R., 1983. The definition of intelligence and factorscore indeterminacy. Behavioral and Brain Sciences 6, 313–315. doi:10.1017/S0140525X00016137.

Kaiser, H.F., 1960. The Application of Electronic Computers to Factor Analysis. Educational and Psychological Measurement 20, 141–151. doi:10.1177/001316446002000116.

Kaiser, H.F., 1974. An index of factorial simplicity. Psychometrika 39, 31–36. doi:10.1007/BF02291575.

Kline, P., 1999. The handbook of psychological testing routledge. Second ed., Routledge, London.

Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., Adesope, O., 2015a. Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. The Internet and Higher Education 27, 74–89. doi:10.1016/j.iheduc.2015.06.002.

Kovanović, V., Joksimović, S., Gašević, D., Siemens, G., Hatala, M., 2015b. What public media reveals about MOOCs: A systematic analysis of news reports. British Journal of Educational Technology 46, 510–527. doi:10.1111/bjet.12277.

Kozan, K., Richardson, J.C., 2014a. Interrelationships between and among social, teaching, and cognitive presence. The Internet and Higher Education 21, 68–73. doi:10.1016/j.iheduc.2013.10.007.

Kozan, K., Richardson, J.C., 2014b. New exploratory and confirmatory factor analysis insights into the community of inquiry survey. The Internet and Higher Education 23, 39–47. doi:10.1016/j.iheduc.2014.06.002.

Lawrence, F.R., Hancock, G.R., 1999. Conditions Affecting Integrity of a Factor Solution under Varying Degrees of Overextraction. Educational and Psychological Measurement 59, 549–579. doi:10.1177/00131649921970026.

Matthews, D., 2013. The Tuition is Too Damn High, Part IX: Will MOOCs save us? The Washington Post URL: http://www.washingtonpost.com/blogs/wonkblog/wp/2013/09/05/the-tuition-is-too-damn-high-part-ix-will-moocs-save-us/.

Poquet, O., Kovanović, V., Vries, P.d., Hennis, T., Joksimović, S., Gašević, D., Dawson, S., 2016. Social presence in Massive Open Online Courses. Manuscript submitted for publication .

Rourke, L., Anderson, T., 2004. Validity in quantitative content analysis. Educational Technology Research and Development 52, 5–18. doi:10.1007/BF02504769.

Rourke, L., Anderson, T., Garrison, D.R., Archer, W., 1999. Assessing Social Presence In Asynchronous Text-based Computer Conferencing. The Journal of Distance Education 14, 50–71. URL: http://www.ijede.ca/index.php/jde/article/view/153.

Shea, P., Bidjerano, T., 2009. Community of inquiry as a theoretical framework to foster "epistemic engagement" and "cognitive presence" in online education. Computers & Education 52, 543–553. doi:10.1016/j.compedu.2008.10.007.

Shea, P., Sau Li, C., Pickett, A., 2006. A study of teaching presence and student sense of learning community in fully online and web-enhanced college courses. The Internet and Higher Education 9, 175–190. doi:10.1016/j.iheduc.2006.06.005.

Siemens, G., 2013. Massive open online courses: Innovation in education, in: McGreal, R., Kinuthia, W., Marshall, S. (Eds.), Open Educational Resources: Innovation, Research and Practice. Commonwealth of Learning, Vancouver, Canada, pp. 5–15.

Snook, S.C., Gorsuch, R.L., 1989. Component analysis versus common factor analysis: A Monte Carlo study. Psychological Bulletin 106, 148–154. doi:10.1037/0033-2909.106.1.148.

Stacey, P., 2013. The pedagogy of MOOCs. The International Journal for Innovation and Quality in Learning 11, 111–115. URL: http://empower.eadtu.eu/images/fields-of-expertise/OERsMOOCs/INNOQUAL-Issue-3-Publication-Sep-2014-FINAL-w-cover.pdf.

Swan, K., Ice, P., 2010. The community of inquiry framework ten years later: Introduction to the special issue. The Internet and Higher Education 13, 1–4. doi:10.1016/j.iheduc.2009.11.003.

Swan, K., Shea, P., Richardson, J., Ice, P., Garrison, D., Cleveland-Innes, M., Arbaugh, J., 2008. Validating a measurement tool of presence in online communities of inquiry. E-mentor 2, 1–12. URL: http://www.e-mentor.edu.pl/artykul/index/numer/24/id/543.

Tabachnick, B.G., Fidell, L.S., 2007. Using multivariate statistics. Pearson/Allyn & Bacon, Boston, MA.

Tinsley, H.E.A., Kass, R.A., 1979. Factor analysis of the domain of client expectancies about counseling. Journal of Leisure Research , 120–138 doi:10.1037/0022-0167.27.6.561.

Velicer, W.F., 1974. A Comparison of the Stability of Factor Analysis, Principal Component Analysis, and Rescaled Image Analysis. Educational and Psychological Measurement 34, 563–572. doi:10.1177/001316447403400309.

Velicer, W.F., Peacock, A.C., Jackson, D.N., 1982. A Comparison Of Component And Factor Patterns: A Monte Carlo Approach. Multivariate Behavioral Research 17, 371–388. doi:10.1207/s15327906mbr1703_5.

Widaman, K.F., 1990. Bias in pattern loadings represented by common factor analysis and component analysis. Multivariate Behavioral Research 25, 89–95. doi:10.1207/s15327906mbr2501_11.

Widaman, K.F., 1993. Common Factor Analysis Versus Principal Component Analysis: Differential Bias in Representing Model Parameters? Multivariate

Behavioral Research 28, 263–311. doi:10.1207/s15327906mbr2803_1.

Widaman, K.F., 2007. Common factors versus components: Principals and principles, errors and misconceptions, in: Cudeck, R., MacCallum, R.C. (Eds.), Factor analysis at 100: Historical developments and future directions. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 177–203.

Winter, J.C.F.d., Dodou, D., 2016. Common Factor Analysis versus Principal Component Analysis: A Comparison of Loadings by Means of Simulations. Communications in Statistics - Simulation and Computation 45, 299–321. doi:10.1080/03610918.2013.862274.

Zwick, W.R., Velicer, W.F., 1986. Comparison of five rules for determining the number of components to retain. Psychological bulletin 99, 432. doi:10.1037/0033-2909.99.3.432.

6.3 Summary

The CoI model was originally designed for understanding and supporting inquiry-based learning via asynchronous online discussions. However, the flexibility in the definitions of three presences and its compatibility with traditional values of higher education in developing critical discourse and reflection (Arbaugh et al., 2008), resulted in the widespread adoption of the CoI model as a most prominent model of teaching and learning in online and blended learning contexts (Arbaugh et al., 2008; Akyol et al., 2009; Gorsky et al., 2011). Although the CoI model has been widely used in online and blended learning research, its validation within the MOOC context received limited attention. The adoption of the CoI model for assessing students' learning experiences in MOOCs is challenging, especially from the methodological standpoint, given many unique characteristics of MOOCs in comparison to conventional online and blended courses such as 1) the omission of course credits, 2) several orders of magnitude larger student cohorts than in conventional, 3) significantly higher demographic diversity of students, 4) broader range of motivational factors for course participation, 5) heavy use of pre-recorded video lectures, 6) and sporadic use of online discussions. The work presented in this chapter is one of the first attempts at validating the CoI model for assessing students' learning experience in MOOCs. Our results confirm the reliability and validity of the model for assessing learning of MOOC students.

From the standpoint of supporting cognitive presence assessment and, in more general terms, the adoption of the CoI model in MOOC setting, the results of the study, presented in this chapter, provide important insights into the dynamics of the CoI presences within MOOC courses. As indicated by the three additional factor loadings, the MOOC context does provide challenges with regards to the development of high levels of cognitive presence, most notably the resolution phase of the practical inquiry model. Similarly, the extensively large student cohorts put more emphasis on strong course organization and design, given the limited opportunities to alter the course once it starts. Finally, the large student cohorts and typically shorter courses provide fewer opportunities for developing affection among the peer learners which also reflects on the climate in social interactions and discourse.

Given that cognitive presence assessment in MOOCs through learning analytics can be used to monitor students' learning during courses, it can provide instructors with the tools for overcoming the challenges associated with large, diverse students cohorts identified in this chapter. For example, the learning analytics models presented in Chapter three and Chapter four can provide means of assessing the progress of large groups of students, and for directing instructors' attention to students and discourse where it is most required. This has a potential to promote the development of high levels of cognitive presence and to stimulate more productive and active student discussion participation. Finally, although not directly related to course organization and design, the active monitoring of student learning provides opportunities for personalized interventions which target large groups of students who do not exhibit desired learning behavior. In the next chapter, building on the validation of the CoI model presented here, we describe an automated analytics system that provides an assessment of students cognitive presence through the analysis of their trace data records.

7 Assessing cognitive presence within MOOC courses

Education is the most powerful weapon we can use to change the world.

— Nelson Mandela, *Notes to the Future: Words of Wisdom*

7.1 Introduction

The introduction of MOOCs to the landscape of online learning was welcomed with great enthusiasm (Pappano, 2012). With MOOC reaching the unprecedented number of students, they have been seen as a panacea for a broad range of issues, such as increasing access to higher education, student debt crisis, providing means for lifelong learning, and the overall democratization of learning (Kovanović, Joksimović, Gašević, Siemens, & Hatala, 2015b, 2015a). While the early enthusiasm for MOOCs has declined over time (Boxall, 2012), MOOCs still provide significant opportunities for improving the understanding of learning processes (Gašević, Kovanović, Joksimović, & Siemens, 2014; Joksimović et al., 2017), given the vast amounts of data being collected and made available to the researchers (Reich, 2015).

Although MOOCs show a great potential for improving student learning experience and to better understand human learning, there are significant challenges related to the current MOOC pedagogical practices which do not build upon the existing knowledge from educational research (Bates, 2012; Stacey, 2013; Bali, 2014). Due to vast numbers of learners, most MOOCs (so-called x-MOOCs) focus on the instructivist knowledge transmission, with very limited social interactions in the course or direct instructional support to the individual learner (Kalz & Specht, 2013; Rodriguez, 2012). This is in sharp contrast to the contemporary educational research which stresses the importance of social interactions for the productive learning experience (Rourke, 2000; Larkin, 2009; Garrison, 2011). As such, there is a need for understanding how the existing models of online learning widely used for small-scale, for-credit online courses can be utilized at scale and within the context of MOOCs and other similar models of online education (Bates, 2012).

Another important characteristic of MOOCs is that – in many cases – they are not-for-credit courses and do not charge enrollment fees [1], which result in a much greater diversity of the student body (Ho et al., 2014). Particularly important are substantial differences in course enrollment motivations (Kizilcec & Schneider, 2015), with a significant number of students not focusing on the course completion in a traditional sense (e.g., completing all course assignments or obtaining a certificate), which in turn produces low course completion rates (Clow, 2013). Similarly, some courses – such as MOOCs based on connectivist learning theories (Siemens, 2005) – do not focus on formal assessments, and do not contain graded assessments nor completion certificates (Daniel, 2014). As a result, *"a pressing question in current MOOC debates is about how to measure their success and quality,"* (Ross, Sinclair, Knox, Bayne, & Macleod, 2014, p. 64) which is for the moment mostly focused on course completion rates (Sandeen, 2013).

In this thesis, we argue that the use of automated learning analytics models can be used to "scale-up" the existing online learning models such as the CoI model. Not only can this lead to improvements in the current MOOC pedagogical practices, but it also has a potential to lead to more theoretically sound research founded on the existing knowledge from distance and online learning. By providing instructors with support in managing course social interactions, their limited time can be directed towards a part of the discourse where their help is most needed. Similarly, by providing instructors with the information about students' use of available tools and resources, they can devise different instructional interventions and support strategies which will be tailored based on students adopted study strategy. Finally, as the analytics developed in this thesis provide rich insights into student course activities, those analytics enable assessment of students' learning beyond simple final course grades. This type of analytics is especially important for learning contexts such as MOOCs, where the absence of course credits completely changes the dynamics of class participation.

In this chapter, we present the study that examined students' use of available technology and how it relates to their levels of cognitive presence within the MOOC setting. This chapter follows the evaluation of the CoI model presented in Chapter six, which confirmed model's validity for modeling student learning experience within MOOC settings. Through the analysis of the trace data of students' use of the MOOC platform, we developed a learning analytics system which can be utilized to assess cognitive presence development of MOOC students from the personal, reflective side of inquiry-based learning. The foundation of the work in this chapter represents the conceptual model of cognitive presence assessment from Chapter two and the learning analytics system presented in Chapter four. Similar to the learning analytics system described in Chapter four, we developed a clustering system to identify common student learning strategies, and then examine how those strategies relate to students' cognitive presence development. Given that in this study we adopted extensive pre- and post-course surveys, including the CoI survey (Arbaugh et al., 2008), we evaluated the identified learning strategies with respect to students' answers to pre- and post-course surveys, as well as examined their differences in social and teaching presence. Our results reveal

[1] It should be noted that some MOOC providers recently started to charge small fees for course certificates or some parts of the course content (Shah, 2017).

significant differences regarding students' use of the available tools and resources which are also associated with the discrepancies in students' levels of cognitive presence. Finally, differences in technology use were found to be associated with self-reported enrollment motivations, prior knowledge, and self-regulated learning skills as measured by the students' pre- and post-course surveys.

7.2 Publication: The role of technology use on shaping student learning experience in MOOCs

The following section includes the verbatim copy of the following publication:

> Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., de Vries, P., Hatala, M., Dawson, S., Siemens, G., and Gašević, D. (2017). The role of technology use on shaping student learning experience in MOOCs. *Manuscript submitted for publication*

The role of technology use on shaping student learning experience in MOOCs

Vitomir Kovanović[a,*], Srećko Joksimović[a], Oleksandra Poquet[b], Thieme Hennis[c], Pieter de Vries[c], Marek Hatala[d], Shane Dawson[b], George Siemens[e], Dragan Gašević[a]

[a]The University of Edinburgh, Old College, South Bridge, Edinburgh EH8 9YL, United Kingdom
[b]University of South Australia, City West campus 217/243 Hindley St, Adelaide, SA 5000, Australia
[c]Delft University of Technology, Jaffalaan 5, 2628 BX, Delft, The Netherlands
[d]Simon Fraser University, Central City Mall, 250-13450 102nd Avenue, Surrey, BC V3T 0A3, Canada
[e]The University of Texas at Arlington, 246 Nedderman Hall, Arlington, TX 76019-0012, United States

Abstract

This paper presents a study which examined student technology use within a massive open online course and its relation to the student learning experience. The theoretical foundation was given by the community of inquiry (CoI) model of online education, which outlines the three critical dimensions (presences) of student learning experience: teaching, social, and cognitive presence. The perceived levels of the three presences were measured by the CoI survey instrument which was administered as the part of the post-course survey. The results of cluster analysis revealed three different technology use profiles: Disengaged users, Strategic users, and Engaged users, which significantly differ regarding the use of available tools and resources as well as the perceived levels of cognitive presence. The results also revealed the differences regarding commitment to learning, motivations and goals for enrolling in the course, goal orientation, approaches to learning, and the use of different study strategies. Implications for research and practice of online learning are further discussed.

Keywords: Community of inquiry model, Massive open online courses, Online learning, Higher Education

1. Introduction

Arguably, one of the most interesting developments in the domain of online and distance education is the emergence of Massive Open Online Courses (MOOCs). Although modalities of individual MOOCs differ substantially, for the most part, they are freely available, fully online, not-for-credit courses that can be enrolled by anyone interested in the topic. While originally introduced as a "revolution" in higher education (Friedman, 2012), MOOCs are better seen as an evolutionary step in the long history of distance education (Daniel, 2014). However, unlike previous generations of distance and online education, MOOCs provide exciting opportunities for using vast amounts of student-generated data for improving the current instructional approaches as well as for understanding the complexities of human learning (Reich, 2015)

There are several important characteristics of the MOOC context which shape student learning experience. First of all, the number of students is much larger than in a traditional for-credit online or blended courses, often reaching up to tens or even hundreds of thousands of students (Coughlan, 2015). The unprecedented size of students cohorts in MOOCs further challenges the commonly applied pedagogical approaches, making it difficult to scale "traditional" teaching practices to this (relatively) new setting. For this reason, current MOOC pedagogies focus primarily on content transmission, with technology being mainly used to scale the behaviorist models of learning and teaching. Secondly, the student population in MOOCs is much more diverse (Ho et al., 2014), with substantial differences in their prior knowledge, age, education level, or proficiency with the English language to name a few. Likewise, the motivations for enrolling in MOOCs is much more diverse (Kizilcec and Schneider, 2015) than in the formal educational setting. Together, these differences render MOOC learning context more challenging for the development of social interactions (Poquet et al., 2016; Akyol and Garrison, 2008; Garrison, 2011). This is in sharp contrast with the contemporary educational psychology which shows significant benefits of social interactions for the development of essential skills such as critical thinking, creativity, collaboration, and communication[1]. Moreover, modern educational research provides many approaches on how to develop, facilitate, and direct effectively online and blended learning experiences by taking advantage of the modern technological systems for information seeking and knowledge building. In this regard, the research around

*Corresponding author.
Email addresses: v.kovanovic@ed.ac.uk (Vitomir Kovanović), s.joksimovic@ed.ac.uk (Srećko Joksimović), sspoquet@gmail.com (Oleksandra Poquet), thieme@hennis.nl (Thieme Hennis), Pieter.deVries@tudelft.nl (Pieter de Vries), mhatala@sfu.ca (Marek Hatala), Shane.Dawson@unisa.edu.au (Shane Dawson), gsiemens@uta.edu (George Siemens), dragan.gasevic@ed.ac.uk (Dragan Gašević)

[1]Often referred to as 21^{st}-century skills

the widely-used Community of Inquiry (CoI) model (Garrison et al., 1999) provides plenty of empirical evidence on the significance of the synergy of teaching, socialization, and cognition for improving the student learning experience.

It is also important to recognize that successful teaching and learning goes far beyond the provision of different educational tools and resources (Lust et al., 2012). Contemporary educational psychology emphasizes the role of *human agency* in shaping student learning experience and provides substantial evidence on the importance of metacognitive skills to effectively leverage the available technological affordances (Winne, 2006). To improve the overall MOOC learning experience, it is essential to understand ways in which students make use of the available technological affordances and how they affect their learning outcomes. In this regard, several theoretical models, such as self-regulated learning (SRL) (Bjork et al., 2013; Winne and Hadwin, 1998), goal orientation (Senko et al., 2011), and approaches to learning (Trigwell and Prosser, 1991), provide theoretical foundation for understanding students' agency.

This paper reports on the study which investigated how student use of the available tools and resources affected the student overall learning experience in MOOCs. The theoretical foundation for the present study is the Community of Inquiry model, given its holistic view of the online and blended learning experience, and has been recently validated within the MOOC context (Kovanović et al., 2016c). Specifically, here we build on the previous work by (Kovanović et al., 2015) which examined students' technology use within traditional, for-credit online courses. In this paper, we therefore:

1. *Identify different profiles* of students based on the trace data of their technology use, which are indicative of students' agency and decision-making processes surrounding the usage of available study tools.

2. *Examine how these different profiles relate to the student perceptions of cognitive, social, and teaching presence*, the three key dimensions of the Community of Inquiry model, which capture student online learning experience and the development of critical and higher order thinking.

3. *Describe the identified profiles* through the analysis of responses to pre-course and post-course surveys, and their final course grades, which provide supplementary information necessary to better understand the human agency processes shaping the identified technology use profiles.

4. *Evaluate the differences between MOOCs and traditional, for-credit online/blended courses* regarding students' technology use, by comparing the results of the present study with the existing studies on the technology use within traditional online/blended setting.

2. Background work

2.1. Theoretical foundations of student online technology use

2.1.1. The Community of Inquiry model

The Community of Inquiry (CoI) model, by Garrison et al. (1999), is a popular pedagogical model which outlines the critical dimensions that shape student learning experience in online and blended settings. Rooted in the social constructivist notion of learning, the CoI model focuses on knowledge (co)creation through a social interaction among the group of learners (Garrison et al., 1999). The model outlines three key dimensions of learning, also known as presences:

1. *Cognitive presence* focuses on students' development of critical and higher-order thinking through an inquiry-based learning process (Garrison et al., 2001).

2. *Social presence* explains social interactions within the student group and the development of productive social climate, which includes open communication, affective expression, and group cohesion (Rourke et al., 1999).

3. *Teaching presence* outlines instructors' role before and during the course, which includes course organization & design, direct instruction, and facilitation (Anderson et al., 2001).

The central construct within the CoI model is cognitive presence which is formally defined as *"the extent to which the participants in any particular configuration of a community of inquiry are able to construct meaning through sustained communication"* (Garrison et al., 1999, p. 89). It is operationalized through a practical inquiry model, which defines four phases of the inquiry-based learning cycle (Garrison et al., 2001):

1. *Triggering event phase*: In this phase, a particular learning problem or dilemma initiates the learning cycle. In the formal educational context, this is typically done by the instructor but can also be by other students.

2. *Exploration*: During this phase, students explore available learning resources, and brainstorm different ideas relating to a particular learning problem.

3. *Integration*: In this phase, students construct new knowledge by synthesizing the relevant information and discarding and filtering the irrelevant information.

4. *Resolution*: After the new knowledge had been synthesized, students apply it to the problem that started the learning cycle. In the formal learning context, this is typically done through hypothesis testing or vicarious action.

For assessing the levels of three CoI presences, researchers typically employ a qualitative content analysis of the student discussion messages using the pre-defined coding instruments (one for each of the presences). More recently, a self-reported instrument by (Arbaugh et al., 2008) has been developed, and it can be used to assess students' perceived levels of social, teaching, and cognitive presences using 34 Likert-scale questions. Both instruments have been widely adopted, which is reflected in the significant number of validation studies (see Garrison et al., 2010).

Although originally developed for assessing the quality of inquiry-based online learning, the broad applicability of the model resulted in its wider adoption as a general model for assessing students' learning experience in online and blended settings (Anderson and Dron, 2010; Swan and Ice, 2010). In the context of MOOCs, a study by Kovanović et al. (2016c) investigated the use of CoI instrument for assessing the levels of the three CoI presences using the data of 1,487 students from five MOOCs. The results of Kovanović et al. (2016c) study indicated that CoI questionnaire is a valid and reliable instrument for assessing the levels of three presences within the MOOC context. The Kovanović et al. (2016c) study also identified certain unique characteristics of MOOCs and the development of three presences within the MOOC context. With the open nature of MOOC enrolment and large student cohorts, the development of affective expression and higher levels of cognitive presence (in particular resolution) is more challenging. The role of course organization and design is also emphasized, given the somewhat limited potential of instructors to affect student learning experience after the course start (Kovanović et al., 2016c).

2.1.2. Educational psychology views on the technology use

One of the founding principles behind the modern educational research is that *students are cognitive agents who monitor, regulate, and control their learning* (Winne, 2006; Anderson and Dron, 2010). In this regard, self-regulated learning (Winne and Hadwin, 1998; Bjork et al., 2013), a major theory in the contemporary educational psychology, provides the theoretical foundation for understanding the human agency in the learning domain. An important aspect of student self-regulation in online and blended learning settings is that it includes decisions on if, how, and when to use particular digital learning tools and technologies (Azevedo, 2005; Winne, 2006). The previous research indicates that many students fail to regulate their use of the available tools and resources in a way which will maximize their learning, with most tools being significantly underused by the majority of the students (Ellis et al., 2005; Lust et al., 2013a, 2011; Kovanović et al., 2015). This is particularly emphasized in complex online learning environments (Shen et al., 2013), as many students simply lack awareness, knowledge, or motivation to use a particular set of tools (Winne, 2006; Lust et al., 2013a).

With regards to adopted methodologies, one of the primary means of understanding student agency and self-regulation is through students' self-reports on their use of available tools and resources (Winne and Jamieson-Noel, 2002; Zhou and Winne, 2012). While this approach has been widely used (e.g., Lust et al., 2013b; Bliuc et al., 2010; Valle and Duffy, 2009), there are several drawbacks to their use. For example, it was shown that students are not capable of accurately estimating their tool use and were often considerably overestimating time spent using specific tools and resources (Winne and Jamieson-Noel, 2002). Primary reasons for the overestimation of time spent include student subjectivity, and also the incomplete and reconstructed memories of the past events (Zhou and Winne, 2012). Hence, the use of trace data is often recommended (Zhou and Winne, 2012) which are considered more objective (Gonyea, 2005). In this paper, we adopted automated learning analytics methods (Siemens and Gasevic, 2012; Baker and Siemens, 2013) for assessing different ways in which students use the available tools and resources, and then complemented it with the data obtained from pre-course and post-course surveys.

In the context of the CoI model, large number of studies (Akyol and Garrison, 2011; Shea and Bidjerano, 2010, 2012; Shea et al., 2012, 2013, 2014; Garrison and Akyol, 2013) pointed to metacognition and self-regulation as a key for understanding student online learning experience, given its self-directed and social nature (Shea and Bidjerano, 2010). To develop their cognitive presence, students are required to engage in the process of critical thinking, which Paul (2005, p.28) defines as *"the art of thinking about thinking in an intellectually disciplined manner."*. As indicated by Shea et al. (2012, 2013, 2014), due to the both self-reflective and social nature of learning within communities of inquiry, development of cognitive presence requires both self and co-regulation of student learning which are supported through teaching presence activities (i.e., instructional design, direct facilitation, and instruction), as well as through peer guidance.

2.1.3. Major factors affecting student technology use

One of the most significant reasons for the difference in student technology use is their ability to effectively *regulate their learning* activities (Clarebout et al., 2013; Lust et al., 2013a). A study by Lust et al. (2013a) showed that majority of students regulated their use of the available tools, yet only 3% had done it effectively and in accordance with the course objectives. The vast majority of students (59%) used a very limited set of tools, which indicated the lack of ability to effectively regulate learning activities for the given learning tasks (Perkins, 1985). As pointed out by Lust et al. (2012), to fully understand student technology use, it is important to look not only at the activity level, but also at the diversity and consistency of the tool use, which are indicative of the learner's adaptation of the learning strategy in accordance with a particular learning task (Lust et al., 2012; Winne, 1982, 2006; Perkins, 1985).

Another construct that was found to directly affect student's use of the available technology is their *goal orientation* (Lust et al., 2013b). Lust et al. (2013b) found that students who focused on *gaining competence* in a given domain – which are indicative of the mastery goal orientation – had more active use of the available tools and technologies. On the other hand, students who focused on *demonstrating competence* – which is indicative of the performance goal orientation – exhibited a more limited use of the technology (Lust et al., 2013b). It should also be noted that in the more recent studies, the student goal orientation is further specified based on the emotional value given to their standards of performance (i.e., approach vs. avoidance). Thus, mastery-approach goal orientation concerns gaining skills and knowledge while mastery-avoidance goal orientation focuses on avoiding learning failures and skill decline. Similarly, performance-approach goal orientation is mainly concerned with demonstrating performance better than peers, while performance-avoidance on not demonstrating performance worse than peers (Senko et al., 2011).

The theory of *approaches to learning* (Trigwell and Prosser, 1991) is another concept which was shown to be directly related to student use of the available technology. The *deep approaches to learning* – which focus on the understanding of the learning content – are shown to be associated with mastery goal orientation (Phan, 2008) and higher student success (Trigwell and Prosser, 1991). In contrast, *surface approaches to learning* – which focus on the reproduction of the learning content – are shown to be associated with performance goal orientation (Phan, 2008) and lower learning outcomes (Trigwell and Prosser, 1991). A study by Wise et al. (2013) examined the use of student online discussions and identified a connection between the minimal cognitive engagement in online discussions and surface approaches to learning and on the other side, an association between the high cognitive engagement in the discussions and deep approaches to learning. Similarly, the results of (Bliuc et al., 2010) study indicate a connection between the fragmented conception of learning in online discussions (i.e., students see discussions as a mere tool to obtain a correct answer) and surface approaches to learning, and also between the cohesive notion of learning in discussions and deep approaches to learning. However, it is important to acknowledge that previous research showed that strategic behavior could be associated with both deep and surface approaches to learning, depending on the student motivation and goal orientation (Entwistle and Tait, 1990).

2.2. Student technology use profiles

As previously stated, learning in collaborative online and blended learning environments also involves decisions on whether to use a particular learning technology and if so how and when to use it (Azevedo, 2005; Winne, 2006). With this in mind, a large number of studies utilized trace and log data collected by the learning environments to examine different ways in which students use technology and what effect they have on the student learning outcomes. A systematic review by Lust et al. (2012) pointed out that there is a plenty of research evidence which suggests that students differ in the way they use the available technology, as well as that those differences have a substantial effect on students' course performance. With many students failing to effectively use the available technology (Lust et al., 2013a), the common pedagogical claim that the mere provision of the rich toolset is beneficial to learning is severely questioned (Lust et al., 2012).

To identify different profiles of students based on their technology use, most of the studies adopted some form of unsupervised clustering techniques. Some of the popular analysis methods include K-means (Yen and Lee, 2011; Lust et al., 2011, 2013b,a; Kizilcec et al., 2013; Bovo et al., 2013; Agnihotri et al., 2015; Ferguson et al., 2015; Rodrigues et al., 2016), hierarchical clustering (Valle and Duffy, 2009; Wise et al., 2013; Rodrigues et al., 2016; Kovanović et al., 2016b), and model-based clustering (e.g., EM clustering) (Cerezo et al., 2015; Bergner et al., 2015). The interpretation of the identified student profiles is also further guided with several relevant theoretical models, including self-regulated learning (SRL) (Bjork et al., 2013; Winne and Hadwin, 1998), goal orientation (Senko et al., 2011), and approaches to learning (Trigwell and Prosser, 1991).

2.2.1. Commonly reported technology use profiles

Although a large number of studies examined student differences regarding the educational technology use, there are wide variations concerning the number and characteristics of the identified student profiles. Clearly, the details of the adopted analysis procedure (e.g., the list of extracted/collected variables, preprocessing steps, analysis technique/algorithm) have a major impact on the final study findings. Moreover, the characteristics of a particular course (e.g., course design, student population, subject domain) have also been shown to have a major role in shaping the identified student profiles. For instance, Ferguson et al. (2015) analyzed five MOOCs offered by the Open University using the same methodology and observed substantial differences between the courses regarding the identified student profiles. In two of the courses, there were seven profiles identified, while in the remaining three courses there were three, four, and five identified profiles, respectively (Ferguson et al., 2015). Hence, to interpret the results of the present study, we reviewed the existing literature for the common themes and similarities between the published studies. In the rest of this section, we report on some of the common profiles repetitively reported in the literature.

While there have been substantial differences between the published studies, overall, most studies identified three student profiles based on their use of the available technology (Yen and Lee, 2011; Lust et al., 2011, 2013a; Bovo et al., 2013; Agnihotri et al., 2015; Valle and Duffy, 2009; Wise et al., 2013; Rodrigues et al., 2016). The highest number of profiles was reported by Li et al. (2015) who identified nine student profiles based on their interactions with MOOC video lectures.

Among the identified student profiles, the majority of published studies identified a profile of students that characterizes the overall low engagement with the learning system. They also exhibited overall low effort and cognitive engagement in the course, performance goal orientation, surface approach to learning, and poor regulation of their tool use (Wise et al., 2013; Valle and Duffy, 2009; Kovanović et al., 2015; Lust et al., 2011, 2013a,b). Depending on the study, these students were referred to as *"no-users"* (Kovanović et al., 2015; Lust et al., 2011, 2013a,b), *"disengaged users"* (Kizilcec et al., 2013; Bergner et al., 2015; Rodrigues et al., 2016), *"minimalist in effort"* (Valle and Duffy, 2009), *"superficial listeners, intermittent talkers"* (Wise et al., 2013), *"passive participants"* (Milligan et al., 2013; Bovo et al., 2013; Hill, 2013; Sharma et al., 2015), *"inactive"* (Li et al., 2015), or *"low achievers"* (Agnihotri et al., 2015). To keep the terminology consistent, in the rest of this paper, we use the term *"disengaged"* to refer to this profile of technology use.

Another commonly reported of student profiles is characterized by a high level of student engagement and the overall active participation in the course. These students are primarily characterized by a mastery approach goal orientation, high cognitive engagement, deep approach to learning and – not surprisingly – high performance. The labels used for this student profile include *"intensive users"* (Lust et al., 2011, 2013a; Kovanović et al., 2015), *"intensive active users"* (Lust et al., 2013b), *"broad listeners, reflective talkers"* (Wise et al., 2013), *"active participants"* (Hill, 2013; Bovo et al., 2013; Sharma et al., 2015; Milligan et al., 2013), *"high achievers"* (Agnihotri et al., 2015), *"self-driven"* (Valle and Duffy, 2009), *"engaged"* (Rodrigues et al., 2016), *"completing"* (Kizilcec et al., 2013; Bergner et al., 2015), and *"keen completers"* (Ferguson et al., 2015). In the rest of the present study, we use the term *"engaged"* students to refer to students with this particular technology use profile.

The third profile often reported in the literature is defined by a selective use of the available tools and resources. Generally speaking, these students were characterized by the focus on accomplishing a particular learning task and typically exhibit high levels of regulatory behavior (Valle and Duffy, 2009; Kovanović et al., 2015; Wise et al., 2013) and also adopt positive study strategies (Valle and Duffy, 2009; Wise et al., 2013). It is interesting to note that, despite being less mastery approach oriented than intensive users (Lust et al., 2013b), their selective tool use and strategic behavior seems to be more indicative of their efforts to manage their time constraints (Wise et al., 2013) rather their mastery or performance goal orientation. Some of the labels used for this profile include *"selective users"* (Lust et al., 2013a,b), *"task-focused"* (Kovanović et al., 2015), *"get it done"* (Valle and Duffy, 2009), *"Concentrated listeners, integrated talkers"* (Wise et al., 2013), and *"efficiency-oriented"* (Yen and Lee, 2011). In the rest of the study, we used the term *"selective"* students to denote this student group.

2.2.2. Technology use profiles within MOOC setting

It should also be noted that studies looking at the students' technology use within the MOOC context – in addition to the three profiles already described – frequently reported several additional technology use profiles. Those additional profiles are mostly the result of the higher variability in motivational factors behind students' course enrollment (Kizilcec and Schneider, 2015) and variability in the course demographics (Hennis et al., 2016). Due to the limited and focused use of the available tools and resources, these profiles are very similar to "disengaged" profile commonly reported in studies of traditional, for-credit online courses, with the primary difference being related to different student motivation and commitment to learning.

Given the open nature of MOOC enrollment, a significant portion of MOOC students only want to explore and get a sense of a particular topic and do not have the intention of completing graded assignments and obtaining the certificate. Course participation of these students consists primarily of passive interaction with video materials (and optionally in-video quizzes) with very infrequent (if at all) completion of graded assignments and limited course discussion participation. These students are often called *"lurkers"* (Milligan et al., 2013), *"observers"* (Hill, 2013), *"auditing"* (Kizilcec et al., 2013), *"viewer"* (Sharma et al., 2015). It should be noted that similar technology use profiles have also been reported in some studies in blended for-credit courses, where a significant portion of the course is face-to-face and when active participation in online discussions was not mandated. We refer to these students as *"auditing"* students.

Another common profile of students within the MOOC context are students who engage with course video materials for one or two weeks. These students are often called *"samplers"* (Kizilcec et al., 2013; Ferguson et al., 2015; Bergner et al., 2015), *"drop-ins"* (Hill, 2013), *"sporadic"* (Rodrigues et al., 2016), or *"wiki-users"* (Sharma et al., 2015). There are several potential reasons for the behavior of these students. First, given a large number of MOOCs on a similar topic, many students intentionally enroll into several similar courses to get a sense of them before actively committing to one of them. Similarly, many students enroll in courses to find information about a particular topic and do not intend to take an active part with the whole course content. In some sense, these students treat MOOCs as another form of freely available online resources, similar to textbooks. We refer to these students as *"sporadic"* students.

The third common profile of students that is reported in the MOOC context are students who registered for the course but did not actively participate in the course. In most MOOCs, these students represent the majority of course registrants and are commonly referred to as *"enroll only students"* (Kovanović et al., 2016b), *"registered users"* (Dawson et al., 2015), or *"no-shows"* (Hill, 2013). Given that in the present study we only analyzed students who actively participated in the course, we removed this group of students from the analysis.

2.2.3. Technology use within communities of inquiry

Despite recognizing the importance of educational technology, within the context of the community of inquiry model, there have been relatively few studies looking at students' use of the educational technology. Using self-reported instruments, a study by Rubin et al. (2013) examined the association between students' perceptions of the three CoI presences and perceived value of Learning Management System

(LMS) affordances. From the perspective of the present study, of direct importance is the finding that the perceived level of cognitive presence is predicted by the perceived ease of LMS communication and the perceived amount of online information, while the perceived ease of finding information was marginally significant Rubin et al. (2013). These findings were expected, given that effective participation in a community of inquiry and reaching high levels of cognitive presence requires participation in online discussions, as well as engagement in various additional learning activities (e.g., online information-seeking, reading learning materials, assessing knowledge through quizzes, and completing course assignments).

More recently, a study by Kovanović et al. (2015) examined student technology use profiles within the context of traditional for-credit online courses. Kovanović et al. (2015) identified six different technology use profiles: (1) *task-focused users*, which were characterized by the strategic engagement; (2) *content-focused users*, who focused on the use of static course content and resources; (3) *no users*, who had the overall low cognitive engagement; (4) *highly intensive users*, which exhibited the highest level of engagement, both in terms of static resources and discussion participation; 5) *content-focused intensive users*, who have above-average content-related activity; and (6) *socially-focused intensive users*, who exhibit above-average activity in online discussions. The principal finding of the Kovanović et al. (2015) study is that students with different technology use profiles have large differences regarding their levels of cognitive presence and that several profiles were associated with the high levels of cognitive presence. These findings emphasize multiple ways in which students can thrive within communities of inquiry and a need for different instructional approaches to help them to better regulate their use of available tools and resources.

3. Research questions

3.1. Understanding student technology use in MOOCs

Existing literature provides substantial evidence for the differences among the students regarding their technology use and the effect that it has on their learning activities and learning outcomes. However, there is a very limited understanding of technology use within the context of communities of inquiry. To date, only Rubin et al. (2013) and Kovanović et al. (2015) examined students' technology use with respect to the CoI model. As the existing studies focused on traditional, for-credit online courses, the goal of this study is to examine to what extent the patterns of association between the technology use and three CoI presences reported in the literature still hold within the MOOC context. Hence the first research question for the present study is:

> Research Question 1:
> What are the main profiles of student technology use within MOOC communities of inquiry?

Although we expect to find similar technology use profiles as reported in the literature (see Section 2), we are interested in examining how does the open nature as well as the massive scale of MOOCs affects student technology within communities of inquiry. We are also interested in using students' final course grades and responses to course surveys to better understand the identified technology use profiles. Although not the central focus of the present study, the analysis of course surveys and grades would enable to better understand the different technology use profiles with respect to relevant constructs described in Section 2 (i.e., motivation, goal-orientation, self-regulation).

3.2. Understanding the relationship between technology use and communities of inquiry in MOOCs

Besides identifying the main technology use profiles, we are also interested in examining the association between student technology use and the development of three dimensions of the CoI model (i.e., cognitive presence, teaching presence, and social presence). We build on the existing literature on the association between technology use (i.e., Rubin et al., 2013; Kovanović et al., 2015), self-regulation (Garrison and Akyol, 2013; Akyol and Garrison, 2011) and student success within the context of communities of inquiry. Hence, the second important goal of present study is to examine how the effects of internal student learning regulation, agency, and approaches to learning – as manifested through different technology use profiles – affects the perceived levels of three CoI presences. Hence, the second research question of the present study is:

> Research Question 2:
> What is the association between the identified technology use profiles and the perceived levels of cognitive, social, and teaching presence within the MOOC context?

Based on the existing research showing the effect of goal orientation and approaches to learning on the development of deep and critical thinking (Kovanović et al., 2015; Trigwell and Prosser, 1991; Phan, 2008; Entwistle, 2009; Bliuc et al., 2010; Wise et al., 2013; Lust et al., 2013a,b), we expect to find the differences between students in terms of their levels of cognitive presence. Unlike the study by Kovanović et al. (2015) who used quantitative content analysis of 1,787 student discussion messages, in this study we adopted the self-reported instrument by Arbaugh et al. (2008), given the much larger size of our dataset. Thus, in addition to looking at the association between cognitive presence and technology use, the use of self-reported instrument enabled us to also examined the association of technology use with teaching and social presence.

4. Methods and materials

4.1. Study data

The data used in this study comes from the Fall 2014 offer of the "Introduction to Functional Programming" MOOC offered on edX platform by the Delft University of Technology.

The course was eight weeks long, and it focused on introducing basics of functional programming using Haskell programming language. Overall, the course was enrolled by 38,029 students, out of which 23,648 logged into the learning system at least once (for the details, see Hennis et al., 2016). Finally, out of the students who were active in the course, 1,968 obtained the course certificate which required 60% of the final grade. The course was delivered using pre-recorded video lectures and students were expected to complete eleven homework and seven labs, which were implemented as edX multiple-choice questions that each allowed for a single attempt. As answers to each multiple-choice question had to be submitted separately, in total, there were 286 individual submissions for 100% final grade. The course also utilized online discussion as well as a set of wiki pages.

At the start of the course students were administered a pre-course survey which consisted of 60 questions covering basic demographics, motivations for enrolling in the course, expectations for class participation, level of experience with the edX platform, knowledge of course topic, proficiency with English language, and level of support for completing the course (Table A.9). After the course, students were also administered post-course survey (Table A.10) consisting of 97 questions about overall course quality and experience, as well as the use of different course components. The post-course survey also included the 34-item CoI survey questionnaire instrument by Arbaugh et al. (2008) which was used to measure the perceived levels of CoI presences (i.e., teaching presence, cognitive presence, and social presence). The data utilized in the course consisted of typical edX trace logs, discussion data exported in JSON format, and the student answers to the pre-course and post-course surveys. The trace data was available for 23,648 students who performed at least one action in the course, while pre-course and post-course survey data were available for 4,909 and 1,040 students, respectively.

4.2. Measuring instrument

To identify groups of students based on their technology use, we extracted various measures of student course engagement from the course trace data (Table 1). The extracted measures were grouped into seven groups of related measures: 1) course access, 2) assignments, 3) video lectures, 4) course navigation, 5) discussion access, 6) discussion contribution, and 7) discussion reputation. The selected features were based on the work of Kovanović et al. (2015), Lust et al. (2011, 2013a,b), and Valle and Duffy (2009). Unlike in the work of Kovanović et al. (2015), we did not use time-on-task measures, given the much higher diversity in student population (and English language proficiency), and also the challenges related to their extraction (Wise et al., 2013; Valle and Duffy, 2009) and implications for the validity of research findings (Kovanović et al., 2016a). However, given the value of time-on-task measures for capturing student discussion engagement (Kovanović et al., 2015), as a substitute, we extracted the average number of characters used when posting a new discussion, or when posting a comment or a reply. Those two measures together provide additional insight into the level of active engagement in online discussion without the challenges of time-on-task estimation.

To examine the relationship between student technology use and development of three CoI presences, we also extracted ten cumulative measures related to the perceived levels of the sub-components of the three CoI presences. As each of the sub-components was measured by 3-4 questions, we obtained a cumulative measure for each sub-component by averaging the responses to the associated Likert-scale items (from 1: strongly disagree to 5: strongly agree). Overall, we extracted: 1) measures for each of the four phases of cognitive presence (i.e., triggering event phase, exploration phase, integration phase, and resolution phase); 2) three measures for three sub-components of teaching presence (i.e., course organization and design, facilitation, and direct instruction); and 3) three measures for the elements of social presence (i.e., affective expression, open communication, and group cohesion). We also extracted the final course grade and a subset of answers to student pre-course and post-course survey questions which were relevant to present study. In total, we extracted 39 measures from the pre-course survey and 53 measures from the post-course survey which are described in detail in the Appendix (Table A.9 and Table A.10, respectively).

4.3. Data pre-processing

As the edX course data is exported in JSON format, we first imported the edX data into a MongoDB database, which is a popular NoSQL database which also internally uses JSON for representing the data. After the data has been imported, the measures of student course engagement were extracted through a series of MongoDB database queries. Count measures were simply a number of times a particular action (e.g., opening a course wiki) was performed which are commonly used in learning analytics and educational data mining research (Kovanović et al., 2016a). Similarly, the average number of characters used per post or comment were extracted by first querying the database for the list of forum contributions of each student and then calculating the average number of characters used by each student. Finally, before the data was used in the cluster analysis, we standardized all 29 clustering measures (i.e., transformed so that each measure has a mean of zero and the standard deviation of one) as distance-based clustering algorithms, such as K-means clustering and hierarchical clustering, depend on the scale of each of the clustering variables (otherwise, variables with higher magnitude will be rendered more significant in the distance calculation).

4.4. Cluster analysis procedure

The cluster analysis procedure closely followed the approach by Kovanović et al. (2015), given the goal of replicating the study findings within a MOOC context. We used agglomerative hierarchical clustering method using Euclidean distance metric and Ward's agglomeration criteria (Hastie et al., 2013) which were already used by several researchers for the similar

Table 1
Extracted measures used in the present study.

#	Variable	Description
	Trace data measures	
	Course access	
1	`OpenHome`	No. of times course main page was opened.
2	`OpenWiki`	No. of times course wiki was opened.
3	`OpenProgress`	No. of times progress page was opened.
4	`OpenSyllabus`	No. of times syllabus page was opened.
5	`OpenCourseware`	No. of times modules were opened.
6	`OpenCoursewareItem`	No. of times module pages were opened.
	Assignments	
7	`AssignmentStart`	No. of times assignments were opened.
8	`AssignmentSubmit`	No. of times assignments were submitted.
	Video lectures	
9	`VideoLoad`	No. of times video lectures were loaded.
10	`VideoPlay`	No. of times video lectures were played.
11	`VideoPause`	No. of times video lectures were paused.
12	`VideoChangeSpeed`	No. of times video lecture speed was changed.
13	`VideoShowSubs`	No. of times video subtitles were shown.
	Course navigation	
14	`ModuleNext`	No. of times next module link was used.
15	`ModulePrev`	No. of times previous module link was used.
16	`ModuleJump`	No. of times "goto" link was used.
	Discussion access	
17	`ThreadAccess`	No. of times threads were opened regularly.
18	`ThreadAccessInline`	No. of times topics were opened from modules.
19	`DiscussionSearch`	No. of times forum search was performed.
	Discussion contribution	
20	`DiscussionsStarted`	No. of regular topics started.
21	`QuestionsStarted`	No. of QA topics started.
22	`CommentsWritten`	No. of responses/comments written.
23	`ThreadsCharsAvg`	Avg. no. of characters per thread.
24	`CommentsCharsAvg`	Avg. no. of characters per response/comment.
25	`UpvotesGiven`	No. of upvotes given.
26	`CommentsEnd.Given`	No. of comment endorsements given.
	Discussion reputation	
27	`CommentsReceived`	No. of replies/comments received.
28	`UpvotesReceived`	No. of upvotes received.
29	`CommentsEnd.Rec.`	No. of comment endorsements received.
	Outcome measures	
	Course grades	
1	`FinalGrade`	Final course grade.
	CoI: Perceived levels of cognitive presence	
2	`Trig.EventLevel`	Perceived level of triggering event phrase.
3	`ExplorationLevel`	Perceived level of exploration phrase.
4	`IntegrationLevel`	Perceived level of integration phrase.
5	`ResolutionLevel`	Perceived level of resolution phrase.
	CoI: Perceived levels of teaching presence	
6	`Org.DesignLevel`	Perceived level of organization & design.
7	`FacilitationLevel`	Perceived level of facilitation.
8	`DirectInst.Level`	Perceived level of direct instruction.
	CoI: Perceived levels of social presence	
9	`AffectiveExp.Level`	Perceived level of affective expression.
10	`OpenComm.Level`	Perceived level of open communication .
11	`GroupCohesionLevel`	Perceived level of group cohesion.
	Course survey measures	
	Pre-course survey measures	
	See Table A.9.	.
	Post-course survey measures	
	See Table A.10.	.

set of problems (e.g., Wise et al., 2013; Valle and Duffy, 2009). To select the optimal number of clusters, we examined the clustering dendrogram (Fig. 1), and in particular the height of the merging steps between different clusters, which is an indicator of their relative similarity. Finally, after the optimal number of clusters is selected, we summarized the identified clusters by computing cluster centroids (i.e., cluster "center points"), which are calculated as the mean values of all variables of all cluster members.

4.5. Statistical analysis procedure

After the clusters have been identified, we used multivariate analyses of variance (MANOVA) (Tabachnick and Fidell, 2007) to examine the differences among clusters regarding technology use and the perceived levels of cognitive, teaching, and social presence. For assessing the differences between clusters regarding technology use, we conducted a one-way MANOVA using cluster assignment as a single independent variable and 29 measures of technology use as the dependent measures. To check for the differences among clusters regarding their perceived levels of cognitive, teaching, and social presences, we conducted a one-way MANOVA with cluster assignment as a single independent measure and with the measures of presences' sub-components as the dependent measures. Before MANOVA, we checked the homogeneity of covariance assumption using Box's M test and homogeneity of variance using Levene's test. To protect from assumption violations, we used Pillai's trace statistic which is considered to be more robust against assumption violations than more commonly used Wilks' Λ statistic (Field et al., 2012). As a final protection measure, we compared the results of MANOVA with the results of robust rank-based variation of MANOVA (Nath and Pavur, 1985).

In cases where significant MANOVA effect was observed, we followed up with the series of univariate analysis of variance (ANOVA) tests for each of the dependent measures. The use of univariate tests only after a significant multivariate effect was observed is considered a sound protection measure against Type I error rate inflation (Bock, 1985). However, as indicated by Bray and Maxwell (1985), this is only true for dependent variables for which a significant multivariate effect was observed. Hence, to further control for Type I error rate inflation, we used very conservative Bonferroni correction procedure. Similarly to Kovanović et al. (2015), before univariate analyses, we checked the homogeneity of variance using Levene's test, and in cases in which it was significant, we used Kruskal-Wallis test instead. Significant univariate analyses were then followed up with a Tukey pairwise posthoc analysis (or pairwise Kruskal-Wallis comparisons in cases where Levene's test was significant). Finally, as univariate analyses assess the differences in each variable separately, to examine the multivariate differences between the identified clusters we used discriminant factor analysis (DFA), which is another commonly used follow-up technique for significant MANOVA analyses (Field et al., 2012). Combined, ANOVAs and DFA provide a complete and coherent assessment of the differences among the identified groups Field et al. (2012).

To better understand the differences in student technology use, we also compared the identified technology profiles regarding the differences in student final course grades and answers to the pre-course and post-course survey questions (Table A.9 and Table A.10. We first examined the homogeneity of variance in the identified student groups using Levene's test. Depending on the results of this test, used either ANOVA or Kruskal-Wallis test to assess the student difference in their final grades, and survey responses. For few of the questions which used categorical rather than ordinal responses, we used chi-square test of independence to examine the difference in the category distribution between the identified student clusters. To improve the accuracy of the chi-square estimates, as suggested by Field et al. (2012), if a category had the expected value less than one for one of the clusters (i.e., a particular cell of the contingency table), or there more than 20% of the cells had expected values less than five, we either grouped this particular category with some other category or completely removed it from the analysis. For example in pre-course survey question #13, we merged the interest "meeting new friends" with "other," as only a few participants answered with this particular option.

Although in the analysis of student survey responses we performed a large number of comparisons, we opted for not using the popular p-value correction procedures for several reasons. First of all, given that most correction procedures – in addition to lowering the Type-I error rate – also inflate the Type-II error rate (Perneger, 1998), there are many opposing views on if, how, and when they should be used (Bender and Lange, 2001; Gelman et al., 2012; Rothman, 1990; Perneger, 1998; Gordi and Khamis, 2004). The primary concern with correction procedures is that they provide overly conservative corrections that significantly reduce the statistical power of the performed tests. For instance, in the present study, the use of most popular Bonferroni correction in statistical tests on survey data would result in the $p = 0.00054$ cutoff to keep the Type-I error rate at the designated level of $\alpha = 0.05$. The primary reason for overly conservative corrections is that they assume a set of *independent tests* which is typically not the case in practice (Conneely and Boehnke, 2007). Also, they focus on lowering the chance of rejecting the true "global null hypothesis" (a hypothesis stating that all null hypotheses are simultaneously true), which is of limited practical interest for most of the researchers (Rothman, 1990; Bender and Lange, 2001; Perneger, 1998). Although there are resampling-based correction procedures that take into the account the correlations among the multiple tests and consequently provide substantially higher statistical power (see Westfall and Young, 1993; Ge et al., 2003), they are primarily used in genomics research where a single statistical test (e.g., t-test) is extensively performed on the data from the same subjects. In our case, those methods could not be applied as we used different types of tests depending on the type of responses (e.g., chi-square or ANOVA/Kruskal-Wallis)

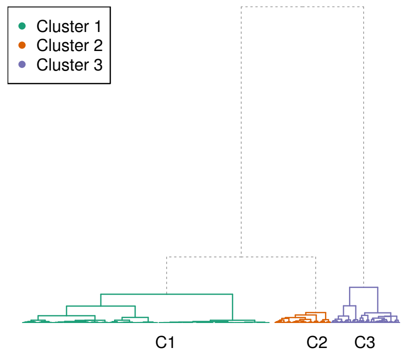


Fig. 1. Dendrogram of student clustering.

and the adequacy of parametric assumptions (e.g., ANOVA or Kruskal-Wallis). Finally, in the present study, the analysis of survey responses is used primarily to provide additional information about the trends and patterns that can aid in the interpretation of the identified clusters. Thus, unlike the analysis of the differences in technology use and three CoI presences, we are not particularly worried about the potentially inflated Type-I error rate as much as the Type-II error rate.

5. Results

5.1. Clustering results

5.1.1. Selecting the optimal number of clusters

The clustering dendrogram is shown in Fig. 1. Our analysis indicated the optimal number of three clusters, one large cluster one with 15,868 students, cluster two with 3,532 students, and cluster three with 4,248 students. The centroids of the identified clusters are shown in Fig. 2, while the detailed scores of clustering variables for the identified clusters are available in the Appendix in Table A.11. Looking at the clustering dendrogram and cluster centroids, we can see that the cluster three is the most distinct cluster, while clusters one and two are more similar to each other, primarily with respect to the use of online discussions. To select the optimal number of clusters, we evaluated all clustering solutions starting with the two-cluster solution. While the two-cluster solution was also viable, the three-cluster solution was preferred given the significant and consistent differences between clusters one and two. The solution with four clusters was also examined, in which case cluster three is divided into two very small and similar clusters. Based on this, we decided to use the three-cluster solution as optimal, and it is the only clustering solution used in the rest of this paper. To enable the easier reporting and interpretation of the results, we assigned each cluster a label (Table 2) based on the interpretation of the key characteristics of each cluster. Section 6 provides an in-depth discussion of the identified clusters and their differences.

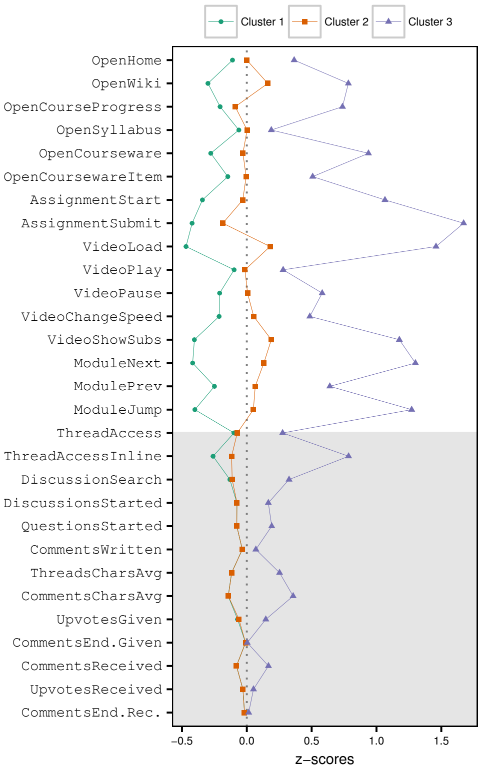


Fig. 2. Cluster analysis results.

5.1.2. *Description of the identified clusters*

Looking at the Fig. 2 and Table A.11, we can see that cluster one students, which we labeled “disengaged users” and which represent two-thirds of the students in the course, have the overall below average engagement in the course, focusing on viewing video lectures, without completing of graded assignments and with very little use of online discussions. In contrast, students from the second cluster, which we labeled “strategic users” and which represent 15% of the student population, show much more engagement with different course material, including course wiki, learning materials, video lectures, and graded assignments. They also show the higher volume of discussion reading activity, while they still do not personally contribute to the discussions. Finally, students from cluster three, which we labeled “engaged users” and which represent 18% of the students in the course, show a strong overall engagement which spans across all course components and activities. In sharp contrast to students from cluster one and two, students from the third cluster show active participation in online discussions which included starting new discussions, posting comments, as well as upvoting and endorsing other posts/comments.

Table 2
Key technology use differences between the identified clusters.

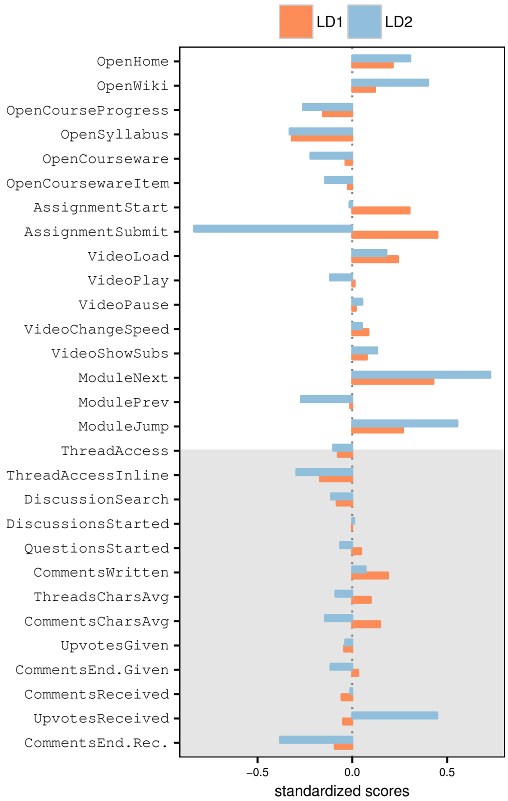
Cluster #	Students (%)	Label	Characteristics
Cluster 1	15,868 (67%)	Disengaged users	Low course engagement, No discussion activity.
Cluster 2	3,532 (15%)	Strategic users	Average course engagement, Almost no discussion activity.
Cluster 3	4,248 (18%)	Engaged users	High course engagement, Use of all course resources
Total	23,648 (100%)		

5.2. *Cluster differences in student technology use*

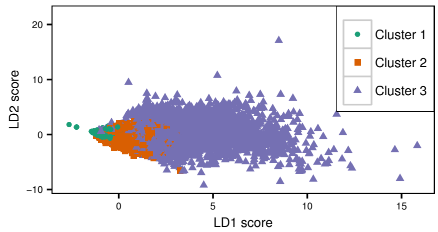
To examine the technology use differences between the identified clusters, we conducted a one-way MANOVA analysis using cluster assignment as a single independent measure, and 29 clustering variables (Table 1) as the dependent measures. Before MANOVA, we used Box’s M test to examine the homogeneity of covariance, and it showed significant differences among three clusters, $p < 0.001$. Hence, we used Pillai’s trace statistic which is shown to be robust against violation assumptions (Field et al., 2012). We obtained significant MANOVA effect, Pillai’s Trace = 0.86, $F(58, 4, 724) = 614.4, p < 0.0001$, with multivariate effect size $\eta^2 = 0.43$ which is considered a large effect size (Cohen, 1988). This indicates that 43% of the variability in the canonically derived dependent variable can be accounted for by the student’s cluster membership. We also used robust rank-based MANOVA (Nath and Pavur, 1985) to confirm our findings. Prior to robust rank-based MANOVA, we removed seven discussion-related variables (i.e., `DiscussionsStarted`, `QuestionsStarted`, `CommentsEndorsementsGiven`, `CommentsEndorsementsReceived`, `CommentsReceived`, `ThreadsCharsAvg`, and `UpvotesReceived`) because of low variability in clusters one and two. We obtained significant MANOVA results, Wilks’ $\Lambda_{rank} = 0.20, \chi^2(44) = 3,7590, p < 0.0001$.

After significant multivariate effects had been observed, we conducted a series of univariate posthoc analyses for each of the dependent variables. Primarily because the low activity of cluster one, Levene’s F test indicated a significant departure from homogeneity of variance for all clustering variables and thus, we used one-way Kruskal-Wallis test with Bonferroni correction. The tests indicated significant differences in regards to all clustering variables, $p < 0.001$. Follow-up pairwise analysis indicated that all three cluster pairs were significant for variables in course access, assignments, video lectures, course navigation, and discussion access groups (Table 3), while for variables in discussion contribution and discussion reputation groups, only the differences between cluster one and three, and two and three were significant.

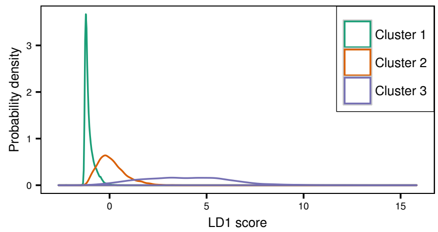
In addition to univariate posthoc analyses, we also used discriminant factor analysis (DFA) to examine the multivariate



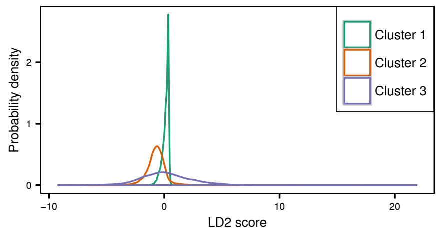
(a) Clustering variable loadings on the two discriminant functions.



(b) Scores for the first two discriminant functions.



(c) The first discriminant function (LD1) score distribution.



(d) The second discriminant function (LD2) score distribution.

Fig. 3. Results of the discriminant function analysis for the multivariate differences between the clusters in terms of the technology use.

differences between three student clusters. Given that we have three clusters in total, DFA produces two linear discriminant functions (i.e., `LD1` and `LD2`), which are linear combinations of dependent measures which explain the most variability in the single independent variable (i.e., cluster assignment). The standardized loadings of two discriminant functions are visualized in Fig. 3a and presented in detail in Appendix in Table A.13. The first discriminant function (`LD1`) explained 97% of the variability in the cluster assignment, while the second discriminant function explained the remaining 3% of the variability. The high score on `LD1` was positively associated with the number of course logins, discussion posting activity, the use of course assignments, wiki, and navigation functionalities (next module and jump to module functionalities), and also the number of video loading events. `LD1` was also negatively associated with the number of times course syllabus and progress pages were accessed. Similarly, we can see that high `LD2` score was positively associated with the number of course logins, use of course wiki, navigation functionalities ("next module" and "jump to module" actions), and the number of received upvotes. It was also most negatively associated with the number of assignment submissions, received endorsements, and use of course progress page, syllabus, and discussions.

If we look at the students' distribution of discriminant function scores (Fig. 3b, Fig. 3c, and Fig. 3d), we can see that three student clusters are reasonably well separated, primarily alongside the first discriminant function (which is expected given its 97% of variance explained). Students from cluster one had highly concentrated scores (around -1 for `LD1` and around zero for `LD2`), while students from clusters two and three had more disperse scores. Students from cluster two had slightly higher scores on `LD1` and slightly lower `LD2` scores that students from cluster one. Finally, students from cluster three had the most disperse scores on both discriminant function, with generally much higher `LD1` scores.

Table 3
Results of the ANOVA analysis for the differences between clusters in terms of the technology use. Significance level $\alpha = 0.0125$ (0.05/4).

#	Variable	Levene's $F(2,23640)$	p	Kruskal-Wallis $\chi^2(2)$	p	Sig. pairs
	Course access					
1	**OpenHome**	200.1	< 0.001	11,830	< 0.001	1-2, 1-3, 2-3
2	**OpenWiki**	2,976	< 0.001	6,351	< 0.001	1-2, 1-3, 2-3
3	**OpenProgress**	1,426	< 0.001	10,620	< 0.001	1-2, 1-3, 2-3
4	**OpenSyllabus**	82.81	< 0.001	5,986	< 0.001	1-2, 1-3, 2-3
5	**OpenCourseware**	1,843	< 0.001	12,300	< 0.001	1-2, 1-3, 2-3
6	**OpenCoursewareItem**	220	< 0.001	14,310	< 0.001	1-2, 1-3, 2-3
	Assignments					
7	**AssignmentStart**	4,078	< 0.001	14,640	< 0.001	1-2, 1-3, 2-3
8	**AssignmentSubmit**	17,770	< 0.001	14,380	< 0.001	1-2, 1-3, 2-3
	Video lectures					
9	**VideoLoad**	6,989	< 0.001	13,030	< 0.001	1-2, 1-3, 2-3
10	**VideoPlay**	231.1	< 0.001	10,010	< 0.001	1-2, 1-3, 2-3
11	**VideoPause**	943.6	< 0.001	10,260	< 0.001	1-2, 1-3, 2-3
12	**VideoChangeSpeed**	1,831	< 0.001	4,352	< 0.001	1-2, 1-3, 2-3
13	**VideoShowSubs**	9,206	< 0.001	9,050	< 0.001	1-2, 1-3, 2-3
	Course navigation					
14	**ModuleNext**	8,459	< 0.001	10,860	< 0.001	1-2, 1-3, 2-3
15	**ModulePrev**	2,521	< 0.001	7,796	< 0.001	1-2, 1-3, 2-3
16	**ModuleJump**	6,671	< 0.001	12,590	< 0.001	1-2, 1-3, 2-3
	Discussion access					
17	**ThreadAccess**	448.1	< 0.001	11,120	< 0.001	1-2, 1-3, 2-3
18	**ThreadAccessInline**	3,652	< 0.001	12,020	< 0.001	1-2, 1-3, 2-3
19	**DiscussionSearch**	947.9	< 0.001	5,976	< 0.001	1-2, 1-3, 2-3
	Discussion contribution					
20	**DiscussionsStarted**	323.8	< 0.001	1,733	< 0.001	1-3, 2-3
21	**QuestionsStarted**	332.5	< 0.001	1,726	< 0.001	1-3, 2-3
22	**CommentsWritten**	66.16	< 0.001	4,012	< 0.001	1-3, 2-3
23	**ThreadsCharsAvg**	791.5	< 0.001	2,872	< 0.001	1-3, 2-3
24	**CommentsCharsAvg**	1,230	< 0.001	4,015	< 0.001	1-3, 2-3
25	**UpvotesGiven**	265.6	< 0.001	3,050	< 0.001	1-3, 2-3
26	**CommentsEndors.Given**	5.251	0.005	348.2	< 0.001	1-3, 2-3
	Discussion reputation					
27	**CommentsReceived**	354.7	< 0.001	2,828	< 0.001	1-3, 2-3
28	**UpvotesReceived**	49.14	< 0.001	1,477	< 0.001	1-3, 2-3
29	**CommentsEndors.Rec.**	17.48	< 0.001	527.7	< 0.001	1-3, 2-3

5.3. *Differences in the perceived levels of the CoI presences*

5.3.1. *Cognitive presence differences*

To check for the differences among the clusters in terms of the perceived levels of cognitive presence, we conducted a one-way MANOVA using a cluster assignment as a single independent measure and four cumulative measures of the perceived levels of cognitive presence as the dependent measures. Before MANOVA, we used Box's M test to check the homogeneity of covariance assumption which was not significant, Box's M = 26.29, $p = .16$. However, given the issue with the Box's M test with uneven group sizes (Tabachnick and Fidell, 2007), we used Pillai's Trace statistic to assess the multivariate differences among the clusters. We obtained statistically significant MANOVA results, Pillai's Trace = 0.02, $F(8, 1626) = 2.05$, $p = .038$. The significant findings were confirmed by the robust rank-based MANOVA, which also yielded significant results (Wilks' $\Lambda_{rank} = 0.98$, $\chi^2(8) = 15.85$, $p = .044$). The multivariate effect size of $\eta^2 = 0.01$ was obtained which is considered a small effect size (Cohen, 1988).

The significant MANOVA was followed with one-way ANOVA for each of the four dependent measures of the perceived levels of cognitive presence (Table 4) with Bonferroni correction. Levene's test indicated that all four dependent measures satisfied the homogeneity of variance criteria (Table 4a). Significant ANOVAs were observed for the differences between clusters regarding the perceived levels of resolution, $F(2, 815) = 5.86$, $p = .003$ with an effect size of $\eta^2 = 0.014$ which is considered a small effect size (Cohen, 1988). Looking at the pairwise comparison between clusters, we see that disengaged users (clusters one) and strategic users (cluster two), significantly differ in terms of the perceived levels of resolution, with later having 0.21 point higher scores (on a 5-point Likert scale). Similarly, we observed statistically significant difference between disengaged users (cluster one) and engaged users (cluster three), with later having 0.20 point higher perceived levels of resolution. The difference between strategic users (cluster two) and engaged users (cluster three) was not significant.

In addition to univariate posthoc analyses, we used discriminant function analysis to assess the multivariate differences among the clusters in terms of the perceived levels of cognitive presence (Fig. 4 and Table A.14). The two discriminant functions (`LD1` and `LD2`) accounted for 94% and 6% of the variability in the student cluster assignment.

Looking at the standardized loadings of discriminant functions (Fig. 4a), we can see that high `LD1` scores are associated with high perceived level of integration, and low perceived level of other three phrases, most notably resolution phase. On the other hand, high `LD2` scores are associated with the low perceived level of the first two phases and high perceived levels of the later two phases. Looking at the scores of both discriminant functions (Fig. 4b), we see a much higher overlap between the three clusters, which is expected given the smaller

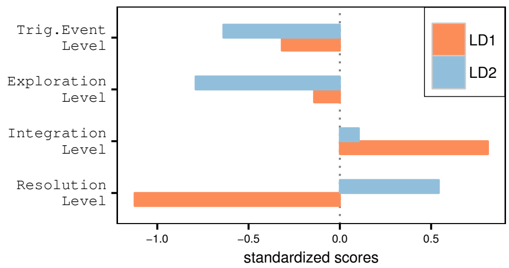
Table 4
Cognitive presence analysis results.

a. Results of the ANOVA analysis for the differences between clusters in terms of the perceived levels of cognitive presence. Significance level $\alpha = 0.0125$ (0.05/4).

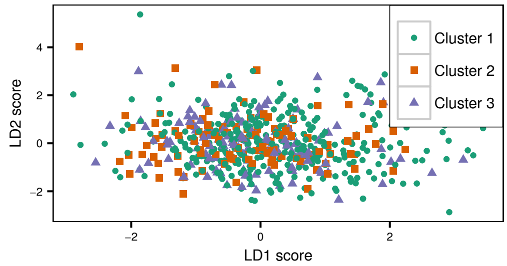
#	Variable	Levene's $F(2,815)$	p	ANOVAs $F(2,815)$	p	η^2
1	`TriggeringEventLevel`	1.46	.23	1.99	.14	.004
2	`ExplorationLevel`	0.66	.52	0.42	.65	.001
3	`IntegrationLevel`	0.42	.65	0.75	.47	.001
4	**ResolutionLevel**	**0.87**	**.42**	**5.86**	**.003**	**.014**

b. Tukey posthoc pairwise comparisons of cluster centers.

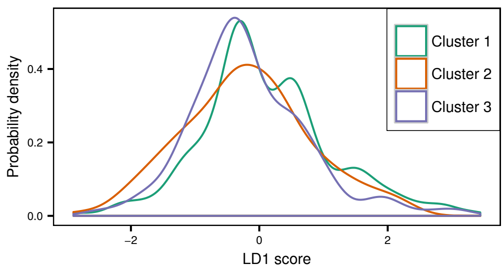
Variable	Cluster Pair	Difference	P adjusted
`ResolutionLevel`	**1-2**	**-0.21**	**0.023**
	1-3	**-0.20**	**0.017**
	2-3	0.01	0.993



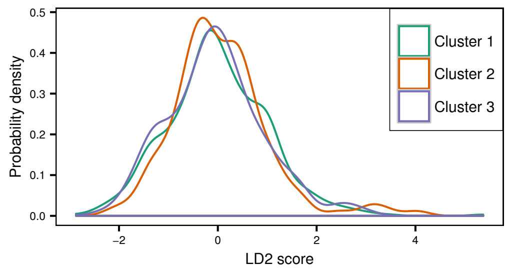
(a) Survey variable loadings on the two discriminant functions.



(b) Scores for the two discriminant functions.



(c) Distribution of the first discriminant function (`LD1`) scores.



(d) Distribution of the second discriminant function (`LD2`) scores.

Fig. 4. Results of the discriminant function analysis for the multivariate differences between clusters in the levels of cognitive presence.

effect size observed. DFA results show that disengaged users had higher `LD1` scores than strategic users and engaged users, whereas strategic users had the most disperse scores, with the highest percentage of students with low `LD1` scores. Looking at the `LD2` scores, we see very little distinction between the three clusters, with strategic users being slightly less likely to have very low `LD2` scores.

5.3.2. Teaching and social presence differences

In addition to examining the cognitive presence differences, we also analyzed cluster differences in the perceived levels of teaching and social presences. To assess the differences in the perceived levels of teaching presence, we conducted a one-way MANOVA using the cluster assignment as the independent variable and three cumulative measures of teaching presence sub-dimensions (i.e., course organization and design, facilitation, and direct instruction) as the dependent measures. To check for the homogeneity of covariance assumption, we used Box's M test which was not significant, Box's M = 12.22 p = .43. As in the case of cognitive presence analysis, given the uneven sizes of our analysis groups, we used Pillai's Trace statistic to check for the multivariate differences among the three clusters. The results of MANOVA indicated no statistically significant differences regarding perceived levels of teaching presence, Pillai's Trace = 0.006, $F(6, 1652)$ = 0.83, p = .54. Our findings were also confirmed with robust rank-based MANOVA, which also produced non-significant findings, Wilks' Λ_{rank} = 0.99, $\chi^2(6)$ = 6.83, p = .33.

To examine the differences in social presence, we adopted the same analysis procedure; We used MANOVA analysis (using the Pillai's trace statistic) with three cumulative measures of social presences sub-dimensions (i.e., affective communication, open communication, and group cohesion) as dependent measures, and cluster assignment as a single independent measure. Box's M test did not indicate a significant departure from the homogeneity of covariances, Box's M = 15.72, p = .21. The MANOVA results indicated no statistical difference between the three identified clusters in terms of their perceived levels of social presence, Pillai's Trace = 0.008, $F(6, 1620)$ = 1.08, p = .37. Finally, same results are confirmed by a robust rank-based MANOVA, which also showed no statistically-significant difference Wilks' Λ_{rank} = 0.99, $\chi^2(6)$ = 4.07, p = .66.

5.4. Analysis of cluster differences in the final course grade

To investigate whether there are significant differences regarding the final course grade between the three clusters of students, we conducted a simple univariate analysis of variance, using the final course grade as a dependent variable, and cluster assignment as an independent variable. However, as significant Levene's test indicated a severe departure from the homogeneity of variance (Levene's $F(2, 23640)$ = 23.29, p < 0.001), we used Kruskal-Wallis test, a corresponding non-parametric alternative. Kruskal-Wallis indicated significant differences among the three clusters in terms of their final course grade, $\chi^2(2)$ = 70.79, p < 0.001. Posthoc pairwise Kruskal-Wallis

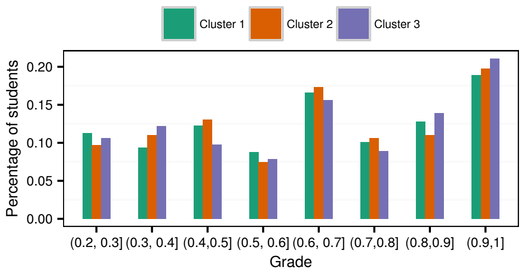


Fig. 5. Percentage of students in each cluster having the final course grade in a particular grade range (grades below 20% are omitted).

comparisons between the three clusters indicated significant differences between cluster 1 (disengaged users) and cluster 3 (engaged users), as well as between cluster 1 and cluster 2 (strategic users), with cluster three having higher mean rank than clusters one and two.

We also examined the distribution of grade scores for students in all three clusters. Figure 5 shows the percentage of students with the course final grade in a given grade range. Given that majority of students with all technology use profiles obtained between 0% and 20% final grade, we omitted them from Fig. 5 to make it easier to see the distribution of students who received higher grades. Overall, we see that the relationship between grade and percentage of students is not linear, which is likely the result of different sub-populations within each cluster and the 60% certificate grade threshold. To investigate this further, we conducted a chi-square independence test using the cluster assignment as a single independent variable, and student grade decile[2] as the dependent measure. We obtained significant results for the difference between in the distribution of grades for the students in three clusters, $\chi^2(18) = 65.63$, $p < 0.000001$. As a posthoc analysis, we looked at the cells in the contingency table (Table 5) with adjusted standardized residuals higher than 1.96, which correspond to the z-score value associated with $\alpha = .05$ (Agresti, 2002). We see that students from cluster one (disengaged users) have statistically significantly higher proportion of 0%–10% grades, and statistically lower proportion of grades in 30%–40%, 60%–70%, 80%-90%, and 90%–100% ranges than what would be expected if there were no differences among the three clusters. In contrast, students from cluster two (strategic users) have statistically lower number of grades in the range 0%-10% range and statistically higher number of grades in the 60%–70% range which is the first range above the certificate threshold of 60%. Finally, cluster three (engaged users) have also statistically lower proportion of grades in 0%–10% range, and higher proportion of grades in the top two ranges (80%–90% and 90%–100%). Interestingly, there is also significantly more engaged users with the grades in 30%-40% range.

[2]Each grade is converted into a value between 0–9 indicating its range (e.g., value 3 indicates that a particular grade is between 30% and 40%).

Table 5
Differences in the final course grade between the identified clusters. Significant cells i.e., those with adjusted standardized residuals (Agresti, 2002) of 1.96 and above, are marked boldface.

Grade range	Value	Cluster 1	Cluster 2	Cluster 3	Grade total
(0%,10%]	Observed:	**13,400**	**2,892**	**3,415**	19,707
	Expected:	**13,223.56**	**2,943.38**	**3,540.06**	
	Adj. std. res.:	**6.55**	**-2.52**	**-5.68**	
(10%,20%]	Observed:	450	104	136	690
	Expected:	463.00	103.06	123.95	
	Adj. std. res.:	-1.07	0.10	1.21	
(20%,30%]	Observed:	227	52	74	353
	Expected:	236.87	52.72	63.41	
	Adj. std. res.:	-1.13	-0.11	1.48	
(30%,40%]	Observed:	**189**	59	**85**	333
	Expected:	**223.45**	49.74	**59.82**	
	Adj. std. res.:	**-4.05**	1.43	**3.62**	
(40%,50%]	Observed:	247	70	68	385
	Expected:	258.34	57.50	69.16	
	Adj. std. res.:	-1.24	1.80	-0.16	
(50%,60%]	Observed:	177	40	55	272
	Expected:	182.51	40.63	48.86	
	Adj. std. res.:	-0.72	-0.11	0.98	
(60%,70%]	Observed:	**335**	**101**	109	545
	Expected:	**365.70**	**81.40**	97.90	
	Adj. std. res.:	**-2.83**	**2.38**	1.25	
(70%,80%]	Observed:	204	51	62	317
	Expected:	212.71	47.35	56.94	
	Adj. std. res.:	-1.05	0.58	0.74	
(80%,90%]	Observed:	**258**	67	**97**	422
	Expected:	**283.17**	63.03	**75.81**	
	Adj. std. res.:	**-2.63**	0.55	**2.71**	
(90%,100%]	Observed:	**381**	96	**147**	624
	Expected:	**418.71**	93.20	**112.09**	
	Adj. std. res.:	**-3.26**	0.32	**3.69**	
	Cluster total:	15,868	3,532	4,248	23,648

5.5. Analysis of cluster differences in course survey responses

We also examined the differences between the identified clusters regarding the answers to pre-course (Table A.9) and post-course surveys (Table A.10). The results of the analysis are shown in Table 6 and Table 8 for pre-course and post-course surveys, respectively. For questions with numerical or Likert scale responses, we first examined the homogeneity of variance assumption using Levene's test and followed with one-way ANOVA or Kruskal-Wallis test, depending on the result of the Levene's test. For a smaller percentage of questions which had categorical responses (e.g., pre-course survey question Q13 regarding the different interests in the course), we used chi-square independence test with cluster assignment as an independent measure.

The results of the analysis of pre-course survey suggest that there is a statistically significant difference between students in the identified clusters regarding the importance of three factors for enrolling into the course: 1) previous experiences with edX,

Table 6
Results of the analysis for the differences between clusters in terms of the responses to pre-course survey questions.

#	Variable	Test	p	Sig. pairs
Basic demographics				
1	`Gender`	$\chi^2(2) = 2.70$	.26	
2	`Age`	$F(2, 3770) = 2.70$	.07	
Enrollment factors				
3	**`ImpFactorPrevEdxExp`**	$H(2) = 19.97$	$< .001$	3–1
4	`ImpFactorPrevTudelftExp`	$F(2, 4005) = 2.07$	.12	
5	**`ImpFactorTudelftRep`**	$H(2) = 7.13$	.03	3–1
6	**`ImpFactorProfRep`**	$F(2, 4017) = 3.87$	.02	3–1
7	`ImpFactorEarningCert`	$F(2, 4022) = 0.05$	.95	
8	`ImpFactorCourseUniq`	$F(2, 4024) = 0.42$	.66	
9	`ImpFactorRecommendation`	$F(2, 4006) = 0.96$	.38	
Motivation				
10	**`DeterminedToLearn`**	$F(2, 4302) = 6.35$	$< .01$	3–1, *2–1*
11	`DeterminedToComplete`	$F(2, 4302) = 1.50$	.22	
12	`Hours`	$F(2, 4300) = 2.70$	.07	
13	**`CourseInterest`**	$\chi^2(6) = 15.26$	.02	see Table 7.
Background knowledge				
14	`LevelOfEd`	$\chi^2(12) = 11.25$	.51	
15	`EdBackFunctProgRel`	$\chi^2(6) = 2.28$	.89	
16	`PrevExpInTheField`	$\chi^2(6) = 11.47$	.07	
17	`Occupation`	$\chi^2(10) = 13.96$	.17	
18	`YearsOfWorkExp`	$F(2, 2700) = 1.17$	.31	
Course expectations				
19	`ExpectFunChallenge`	$F(2, 4297) = 1.66$	.19	
20	**`ExpectRelevance`**	$F(2, 4297) = 3.82$	.02	*3–1*
21	`ExpectForumFeedbackTeam`	$F(2, 4290) = 0.62$	.54	
22	`ExpectForumPeerInteract`	$F(2, 4292) = 0.08$	.91	
Study strategies				
23	**`HowLikelyStudyGroup`**	$F(2, 4294) = 4.20$	.01	3–1
24	`HowLikelyMakeFrieds`	$F(2, 4293) = 0.75$	.47	
25	`HowLikelyLookExtraMat`	$F(2, 4296) = 0.50$	.60	
26	`HowLikelyPostDisc`	$F(2, 4293) = 2.19$	.11	
27	`IntentShareExpertise`	$F(2, 3969) = 0.94$	.39	
28	`PrefAloneVsWithOthers`	$\chi^2(2) = 3.06$	.22	
29	`PrefTeacherVsStudentResp`	$F(2, 4017) = 2.17$	.11	
Study support				
30	`SupportBy`	$\chi^2(2) = 0.18$	.90	
31	`SupportHow`	$\chi^2(2) = 3.09$	.80	
32	`StudyDuringWorkHrs`	$\chi^2(2) = 1.27$	.53	
Experience with online environments				
33	`FreqSocialMediaUse`	$H(2) = 1.75$	.41	
34	`FreqForumUse`	$F(2, 3979) = 0.69$	.50	
35	`OnlineClassesTaken`	$H(2) = 3.67$	.15	
36	`OnlineClassesComp`	$F(2, 3442) = 2.05$	.12	
Language fluency				
37	`EnglishFluencyLevel`	$F(2, 3836) = 0.86$	.42	
38	`EnglishHowComfortable`	$F(2, 3836) = 2.27$	.10	
39	`EnglishHowOften`	$F(2, 3836) = 2.03$	.13	

2) reputation of the Delft University of Technology, and 3) instructors' reputation. The students from cluster three (engaged users) reported a significantly higher importance of these three factors than students from cluster one (disengaged users).

Our analysis also revealed the statistically significant differences regarding the primary interest in the course (Table 7). Students from cluster one were significantly more likely to be driven by the general curiosity in the course subject, while in contrast, students from cluster two and three were significantly less likely to be motivated by the broad interest in the course. In addition, students from cluster three were significantly more driven by the current professional occupation. The ambition to get a degree in course subject domain was not a statistically significant for any of the clusters.

Table 7
Contingency table for the differences in primary motivations for enrolling into the course between the identified clusters. Significant cells i.e., those with adjusted standardized residuals (Agresti, 2002) of 1.96 and above, are marked boldface.

#	Course interest	Value	Cluster 1	Cluster 2	Cluster 3	Interest total
1	*Ambition to get a degree*	Observed:	176	52	56	284
		Expected:	183.66	45.42	54.92	
		Adj. std. res.:	-0.99	1.10	0.17	
2	*Current occupation*	Observed:	**696**	197	**250**	1,143
		Expected:	**739.17**	182.81	**221.01**	
		Adj. std. res.:	**-3.13**	1.34	**2.55**	
3	*Curiosity in the topic*	Observed:	**1,776**	**401**	**489**	2,666
		Expected:	**1,724.09**	**426.41**	**515.51**	
		Adj. std. res.:	**3.49**	**-2.23**	**-2.16**	
4	*Other*	Observed:	61	20	15	96
		Expected:	62.08	15.35	18.56	
		Adj. std. res.:	-0.23	1.31	-0.93	
		Cluster total:	2,709	670	810	4,189

The pre-course survey analysis also revealed a statistically significant difference between students regarding their determination to learn the course materials, with students from cluster three being significantly more determined to learn than students from cluster one (the difference between clusters two (strategic users) and one was marginal, $p = .055$). Although results revealed the differences regarding the course relevance, the follow-up posthoc analysis was not significant (marginal differences between clusters three and one, $p = .061$). Results also revealed the statistically significant differences regarding the use of study groups, with students from cluster three being significantly more likely to join study group than students from cluster one.

Similarly to the analysis of pre-course survey responses, we analyzed student responses to the post-course survey. The results indicate the significant differences between students regarding the use of additional learning materials, with students from clusters three (engaged users) and two (strategic users) reporting higher use of additional resources than students from cluster one (disengaged users). Consistent with the cluster analysis, post-course survey analysis also suggests the difference regarding the use of online discussions, with students from cluster three reported sharing their knowledge significantly more than cluster one students. Interestingly, when asked if they indented to share their knowledge more than they actually did, cluster one students reported significantly higher scores than cluster three students. This indicates that cluster one students had strong intentions for the discussion participation which they fail to realize during the course.

Table 8
Results of analysis for the differences between clusters in terms of the responses to post-course survey questions.

#	Variable	Statistical test	p	Sig. pairs
Course engagement level				
1	`ParticipationLevel`	$\chi^2(6) = 5.47$	.48	
2	`HoursDedicated`	$F(2, 856) = 0.20$	.82	
3	`WorkedHard`	$F(2, 925) = 0.18$	.83	
4	`FamilyWorkObligationsAside`	$F(2, 924) = 0.25$	.78	
5	`PlannedAndOrganizedLearning`	$F(2, 923) = 0.08$	.93	
6	**`LookedAtExtraMaterials`**	$F(2, 095) = 4.19$	.01	3–1, 3–2
Social interactions				
7	`ForumHowOften`	$F(2, 908) = 0.67$	.51	
8	`PostedInDiscussions`	$F(2, 904) = 0.75$	.47	
9	**`SharedExpertiseWithOthers`**	$F(2, 906) = 3.72$	.02	3–1
10	**`IntendedToShareMore`**	$H(2) = 14$	< .001	1–3
11	`ConnWithCourseTeam`	$F(2, 904) = 1.03$	.36	
12	`ConnWithOtherStudents`	$F(2, 906) = 0.77$	.46	
13	`ConnWithNewFriends`	$F(2, 093) = 0.53$	.59	
14	`WouldLikedConnectMore`	$F(2, 905) = 2.95$	.053	
15	`PartInStudyGroup`	$F(2, 903) = 0.30$	.74	
Challenges to course participation				
16	`PersonalMedicalIssue`	$F(2, 954) = 1.18$	.31	
17	`ProfessionalObligation`	$F(2, 956) = 0.03$	.97	
18	`NotGettingFeedback`	$F(2, 955) = 0.08$	.92	
19	`FeelingLonely`	$F(2, 954) = 0.11$	.90	
20	`TechOrAccessibilityIssue`	$F(2, 955) = 0.78$	.46	
21	`KeepingUpWithCoursePace`	$F(2, 954) = 1.63$	.20	
22	`FamilyObligation`	$F(2, 956) = 0.13$	.88	
23	`AnotherCourseObligation`	$F(2, 956) = 0.10$	.90	
Technical challenges to course participation				
24	`SlowInternet`	$F(2, 944) = 0.03$	.97	
25	`InternetNetworkProblems`	$F(2, 944) = 0.52$	.60	
26	`ElectricityNetworkProblems`	$F(2, 944) = 2.12$	.12	
27	`NotHavingAccessPers.Comp.`	$F(2, 944) = 1.13$	.32	
28	`HardwareProblems`	$F(2, 943) = 1.27$	.28	
29	`NotHavingMobileAccess`	$F(2, 945) = 1.53$	.22	
30	`AccessibilityUsability`	$F(2, 944) = 2.50$	.08	
Course appropriate				
31	`CourseExpectationsRealistic`	$F(2, 927) = 1.86$	.16	
32	`CourseMeetYourExpectations`	$F(2, 813) = 1.78$	.17	
33	`LevelOfEnglishAppropriate`	$F(2, 929) = 0.83$	.43	
34	`CourseRelevantToOccupation`	$F(2, 928) = 1.52$	.22	
35	`PriorKnowledgeMadeItEasier`	$F(2, 929) = 0.76$	.47	
Course support				
36	`CourseInspiredToStudy`	$F(2, 895) = 1.45$	.23	
37	`ForumWasHelpfulForMe`	$F(2, 907) = 0.19$	.82	
38	`OthersHelpedMeInTheCourse`	$F(2, 906) = 1.35$	.26	
39	`ReceivedSupportStudents`	$F(2, 917) = 1.29$	.27	
40	`ReceivedSupportCourseTeam`	$F(2, 916) = 1.17$	.31	
41	`SupportFromCourseTeamForums`	$F(2, 862) = 0.25$	.78	
42	`OthersCouldHelpMore`	$F(2, 904) = 2.44$	.09	
43	`ForumCouldBeMoreHelpful`	$F(2, 905) = 2.81$	.06	
Course design and quality evaluation				
44	`DifficultyLevelOfTheCourse`	$F(2, 883) = 1.11$	.33	
45	`AmountOfWorkRequired`	$F(2, 896) = 2.13$	.12	
46	`PaceOfTheCourse`	$F(2, 896) = 0.01$	.99	
47	`DurationOfTheCourse`	$F(2, 897) = 0.26$	.77	
48	`LecturesExercisesBalance`	$F(2, 894) = 1.55$	.21	
49	`CourseOverallQuality`	$F(2, 897) = 0.35$	.70	
50	`AssignmentAndExamQuality`	$F(2, 896) = 0.66$	.52	
51	`VideoLecturesQuality`	$F(2, 895) = 2.08$	.12	
52	`FeedbackQuality`	$F(2, 880) = 1.97$	.14	
53	`EdxEaseOfUseAndQuality`	$F(2, 897) = 1.46$	.23	

6. Discussion

6.1. RQ 1: Technology use profiles withing the MOOC communities of inquiry

The clustering results revealed large differences ($\eta^2 = 0.43$) between MOOCs and traditional online courses regarding students' use of available technology. The observed differences are well aligned with the existing research of student technology use in both traditional online courses (Kovanović et al., 2015; Lust et al., 2011, 2013a,b; Valle and Duffy, 2009) and MOOCs (Kizilcec et al., 2013; Ferguson et al., 2015; Hill, 2013; Rodrigues et al., 2016; Sharma et al., 2015; Milligan et al., 2013; Li et al., 2015). The differences are expected, given the different internal and external conditions upon which students regulate their learning (Winne and Hadwin, 1998; Butler and Winne, 1995), as well as different motivation (Kizilcec and Schneider, 2015), demographics and prior knowledge (Hennis et al., 2016).

Looking at the description of cluster centroids (Fig. 2 and Table A.11), we see a consistent pattern of below-average tool use by the students from cluster one (disengaged users), average use of available technology and resources by cluster two students (strategic users), and active technology use by cluster three students (engaged users). The biggest differences were regarding the use of graded assignments, which is consistent with the several studies that identified differences between MOOC participants in their commitment to course assessment (Hill, 2013; Kizilcec et al., 2013; Sharma et al., 2015; Ferguson et al., 2015). On average, cluster one students submitted one, cluster two students 21, and cluster three students 180 graded submissions (out of total 286 assignments). These differences indicate the strong commitment of students from cluster three of obtaining high course grades and are also consistent with the results of the differences in student final grade (Table 5). We can see that disproportionately many students from cluster two obtained a passing grade (i.e., between 60% and 70%), whereas a considerable number of students from cluster three achieved the highest grades (i.e., between 80% and 100%).

Considering the differences in student use of online discussions, we also see a similar pattern of different engagement of students from the identified clusters. The students from cluster one almost entirely ignored online discussions, cluster two students accessed discussions but did not actively contributed to them, whereas students from cluster three had very active and committed participation in online forums. This is aligned with the findings of Wise et al. (2013) and Kovanović et al. (2015) who identified significant differences in student online discussion participation. Those differences are likely related to different student motivations (Table 6), as well as different conceptions of learning in discussions (Bliuc et al., 2010), and metacognitive monitoring and control (Winne, 2006).

Although we observed differences in the use of both static resources and discussions, the differences in the use of online discussion use were more modest than the differences pertaining to course static resources (Table A.11). Unlike Kovanović et al. (2015) who reported larger effects concerning the use of online discussions, our results show the opposite trend, with larger differences being associated with the use of static course resources. The likely cause for the observed differences regarding the use of online discussions is more careful and detailed planning and scaffolding of student discussion participation in the course examined by Kovanović et al. (2015), which is shown to be directly related to the quality of student use of discussion forums (Gašević et al., 2015).

Examining the variable loadings for the two discriminant functions (Fig. 3a and Table A.13), we see that almost all variability in student cluster membership (97%) is explained by the first discriminant function (LD1) while the second discriminant function (LD2) explained the remaining 3%. From the distribution of student LD1 (Fig. 3c) and LD2 (Fig. 3d) scores, we see that students in the first cluster (disengaged users) have very similar scores on both discriminant functions suggesting very little variability between them. In contrast, students from cluster three (engaged users) who were most active in the course show much higher variability in their scores, while students from cluster two (strategic users) were somewhere between these two extremes.

Looking at the LD1 coefficients, we can see that students primarily differed in the use of course assignments, video lectures, module navigation actions (in particular “next” and “jump” actions), discussion forums, as well as the overall course access (indicated by the `OpenHome` action). Hence, LD1 could be best described as students’ the *overall course engagement*, given the highly positive coefficients for all of the essential course components (i.e., videos, assessment, discussions) and positive learning strategies (e.g, active discussion participation, revisiting behavior, self-directed course navigation, and frequency of course access). These findings are aligned in part with results reported by Kovanović et al. (2015), who also identified the first discriminant function to be representative primarily of the overall course engagement. However, unlike the Kovanović et al. (2015) study results, the LD1 function in the present study accounted for a much higher percentage of variability in student cluster assignment (97% vs. 69%). This is likely caused by more diverse motivational factors within the MOOC context (Kizilcec et al., 2013; Kizilcec and Schneider, 2015) than in the formal, for-credit, online setting.

The coefficients of LD2 were most strongly positively associated with the course overall access, wiki access, navigation functionalities (“next” and “jump”), and the number of upvotes received. It was also strongly negatively associated with assignment submissions, access to progress, syllabus, and the number of comment endorsements received. As such, it is best described as *selectivity of course experience*, as students who obtained high LDA2 scores likely viewed the course as a set of resources and materials. They accessed video lectures in a selective manner (indicated by the positive coefficient for “jump module” action), however, without submitting graded assignments, and accessing the pages related to course progress. The pattern of association with online discussions is challenging to interpret as there are several conflicting coefficients indicating active (`UpvotesReceived`) and passive (`CommentsCharAvg`, `CommentsEnd.Rec`) participation. The insights from the LDA2 are also aligned with the number of studies in the MOOC context, which reported a cluster of selective and strategic MOOC learners (Ferguson et al., 2015; Kizilcec et al., 2013; Hill, 2013; Rodrigues et al., 2016; Sharma et al., 2015).

6.2. *RQ 2: Effects of technology use on the development of the three CoI presences*

Based on the results of the multivariate analysis we see that there were significant differences between the identified clusters regarding the perceived levels of cognitive presence, which is aligned with the findings of Kovanović et al. (2015) in the traditional, for-credit courses. However, unlike the Kovanović et al. (2015) results which show large effect size regarding the cognitive presence differences between the identified clusters, in the present study the multivariate effect size of the differences in the (perceived) levels of cognitive presence were smaller.

There are several likely explanations for the smaller observed differences regarding cognitive presence between the identified clusters. Firstly, given the open nature of MOOCs, the course adopted less structured organization without clear expectations regarding the discussion participation, which might affect the level at which student develop their cognitive presence. Secondly, as with any survey instrument, there is a strong self-selection bias of students who completed the questionnaire. This is visible by the number of students answering the course surveys, which is around 16% for the pre-course survey (Table A.15) and around 4% for the post course survey(Table A.16). Hence, it is likely that higher percentage of active and engaged students from all three clusters are completing the course surveys, rendering the differences between the clusters smaller than they actually are. Finally, given the distinct internal and external conditions (Winne, 2006; Winne and Hadwin, 1998; Butler and Winne, 1995) between students from different clusters arising from the differences in background knowledge, motivation, and course expectations (Kizilcec et al., 2013; Kizilcec and Schneider, 2015), it is likely that students applied different standards of performance when answering the survey questions. For example, when answering the CoI survey question #1 “Problems posed increased my interest in course issues” assessing the student triggering event phase, it is likely that students from three clusters interpreted it differently given the differences in motivations and course expectations between them.

The results of the univariate posthoc analysis indicate that there are significant differences between the three clusters regarding the perceived levels of resolution phase. In particular, students from clusters one (disengaged users) have lower perceived levels of resolution than students from cluster two (strategic users) and three (engaged users). This is aligned with the previous research which pointed to the challenges of reaching higher levels of cognitive presence, in particular, the resolution phase, in the traditional small-scale online courses (Garrison et al., 2001, 2010). Similarly, for students to reach higher levels of cognitive presence, the use of online discussions should be better motivated by the course design (Rovai, 2007; Penny and Murphy, 2009), which includes their planning and adequate scaffolding (Gašević et al., 2015). Finally, these results are aligned with the results of exploratory factor analysis by Kovanović et al. (2016c) which showed survey items related to resolution phase loading on a separate latent factor than the rest of the cognitive presence survey items. These results suggested the different dynamics of reaching resolution phase from the remainder of cognitive presence items, likely due to the shortened course durations, limited role of instructors, and more diverse student populations (Kovanović et al., 2016c).

The results of the multivariate posthoc analysis (Fig. 4 and Table A.14), show that almost all of the variability in student cluster assignment (94%) is explained by the first discriminant function (LD1). Interestingly, all phases except integration have negative LD1 coefficients which mean that students with high LD1 scores exhibited higher levels of integration, relative to other three phases. This is in contrast to the results of Kovanović et al. (2015) study in which all LD1 coefficients had the same direction. We see that students from clusters one had slightly higher scores on LD1 than students from clusters two and three. This indicates that cluster one students exhibited a relatively higher emphasis on integration phase than on the other three phases, whereas students from clusters two and three showed the opposite trend, with the primary focus on the resolution phase. As in the case of univariate analyses, it is likely that differences between students regarding their background knowledge, motivation, and course expectations rendered their use of the available tools and resources in a way which promoted course participation focusing on the different levels of cognitive presence. With students from clusters two and three having a predominantly professional-related interest in the course, it is not surprising to see a higher emphasis on the resolution phase. Similarly, the focus on the integration phase is reasonably aligned with the general interest in the course topic expressed by the cluster one students.

In addition to the analysis of cognitive presence differences among the identified clusters, we also examined their differences with respect to the teaching and social presences. Surprisingly, our analysis did not reveal significant differences in the perceived levels of teaching and social presences between the three identified clusters. While more research is necessary to fully understand the determinants of teaching and social presence in MOOC settings, there are several potential causes for the lack of differences among the three clusters which relate to the unique characteristics of learning within MOOC context and adopted measurement instrument. Firstly, due to the voluntary use of the self-reported instrument, the results of the present study are most likely affected by the self-selection bias. Given that more active and engaged students (from all three clusters) are more likely to fill-in the post-course CoI questionnaire, the cluster differences are rendered smaller than in reality. Secondly, due to the massive number of students, most MOOCs employ pre-recorded video lectures, automatic assessment mechanisms, and very standardized course designs with very little changes during the course. The facilitation and direct instruction of student discourse by the course instructors is also reduced due to the massive volume of interactions in online discussions. As a result, most students are exposed to the very similar course experience, which will be reflected in their perceived levels of teaching presence. Finally, with regards to social presence, previous research pointed to the critical importance of cohort size (Akyol and Garrison, 2008; Garrison, 2011; Poquet et al., 2016) and course duration (Poquet et al., 2016) on the development of social presence, with significant challenges associated with the larger student groups and shorter courses. As the MOOC in present study was eight weeks long and with more than 20,000 students, it is likely that majority of students were not able to develop their social presence to the desired levels.

6.3. The interpretation of the identified technology use profiles

In this section, we provide a detailed interpretation of the identified technology use profiles based on the observed differences in cognitive presence, course final grades, and survey responses. We also build on the existing literature on CoI model, technology use, and important factors that shape students' technology use (e.g., background knowledge, motivation, self-regulated learning, goal orientation), described in detail in Section 2.

6.3.1. Cluster one: Disengaged users

With 15,868 students, cluster one is by far the largest in size, accounting for two-thirds of the student population. Overall, the students from this cluster exhibit the below-average course engagement and low use of the available tools and resources (Fig. 2 and Table A.11). Given their limited use of the available technology, they are most similar to the "disengaged" students reported in the previous studies (see Section 2). They have significantly lower scores on all engagement metrics than students from cluster three (engaged users) and lower scores on all engagement metrics than students from cluster two, except for the metrics related to active discussion participation (Table 3). Students from cluster one also have very similar behavior to one another, which is indicated by their very concentrated LD1 and LD2 scores (Fig. 3c and Fig. 3d, respectively), and smallest standard deviations of all trace data metrics (Table A.11).

Their course participation primarily consists of passive consumption of course's video materials, without completion of graded assignments. Not surprisingly, the majority of cluster one students obtained a final course grade in 0–10% range (Table 5). As such, they are also similar to "auditing" students reported in the studies of technology use within the MOOC context. Based on the analysis of survey responses we see that cluster one students are primarily motivated by the general interest in the course and are significantly less likely to have profession-related interests for participation (Table 7). When deciding whether to enroll the course, previous edX experience and university's and instructors' reputation were significantly less important than for cluster three students. They also did not expect the course to be highly relevant to their current occupation, and as such, have lower determination to learn than students from other two clusters (Table 6). Finally, they were also less likely to use additional learning materials and join study groups (Table 8).

There are several plausible explanations for the behavior of cluster one students. As the way in which students learn online indicates their commitment to learning (Valle and Duffy, 2009), it is likely that students from cluster one lack intrinsic motivation to engage in learning. This is suggested by their lower determination to learn, look extra materials, participate in study groups, and by the lack of keen interests in the course content. Given their very limited engagement in active learning activities, it is also likely that cluster one students have surface approaches to learning and performance-avoidance goal orientation (Phan, 2008).

With respect to the development of cognitive presence, students from cluster one reported lower perceived levels of resolution phase than students from other two groups (Table 4b). This is likely caused by the lack of work-related interests in the course which would provide students with incentives to further develop their cognitive presence. Another contributing factor is their focus on content-related activities, without active participation in online discussions. Interestingly, while they did not participate in the forums, they reported having intentions to participate in online discussions. Hence, it is likely that these students have poor self-regulating skills required to successfully participate in online discussions (Hew et al., 2010), and is consistent with the research showing the challenges with learning regulation by a significant portion of student population (Dunlosky and Lipko, 2007; Lust et al., 2013a; Kovanović et al., 2015). As such, these students might benefit from instructional support regarding the interactions with other students (Cho and Kim, 2013).

6.3.2. Cluster two: Strategic users

Cluster two encompasses 15% of students in the course, which are characterized by the average use of the available technology, with a particular focus on course video materials and graded assignments (Fig. 2 and Table A.11). Like cluster one students, they also do not show active participation in online discussions and use them primarily as a static learning resource. Students from cluster two also show much higher variability in their behavior, which is indicated by the wider distributions of LD1 and LD2 scores (Fig. 3c and Fig. 3d, respectively), as well as larger standard deviations for all trace data measures (Table A.11). Statistical analyses showed that they have significantly higher values for all trace data measures, except for ones relating to active discussion participation (Table 3). Given their selective and goal-oriented use of available tools and resources, they are most similar to the broad class of "selective" and "strategic" students reported by a large number of studies covered in Section 2.

If we look at the grade distribution of students from cluster two, it is clear that these students are primarily focused on obtaining the course certificate, with a higher-than-expected number of students having grade just above the 60% threshold level (Table 5). Compared to cluster one students, our results showed marginally significant difference regarding their dedication to learning (Table 6) and also are less likely to participate in the course due to their general curiosity in the topic (Table 7). As their goal is primarily to obtain the course certificate, they are also less likely to look extra course materials (Table 8). With regards to their development of cognitive presence, they show higher perceived levels of resolution phase than students from cluster one and no differences regarding the levels of cognitive presence with students from cluster three. The greater development of cognitive presence is likely caused by their higher commitment to learning and enrollment interests other than "curiosity in the topic" (which is predominantly reported by cluster one students).

While the purposeful and selective use of different learning tools is an indication of the ability to regulate learning, it can be combined with both surface and deep approaches to learning (Entwistle, 2009). In the case of cluster two students, their highly selective and focused use of tools indicates that these students have likely adopted performance goal orientation (Phan, 2008), coupled with surface approaches (Trigwell and Prosser, 1991) to learning with an ultimate goal of obtaining the course certificate. Moreover, based on their limited and mostly passive use of online discussions, it is likely that these students do not perceive participation in online forums as a valuable learning activity (Dennen, 2008) and likely have fragmented notion of learning in online discussions (Bliuc et al., 2010). As such, similarly to cluster one students, it is likely that instructional interventions which target their interaction with other students (Cho and Kim, 2013). As indicated by Wise et al. (2013), for students who employ performance goal orientation, designing and embedding active participation in the activity requirements can be particularly beneficial and encourage the development of cognitive presence and deeper knowledge comprehension.

6.3.3. Cluster three: Engaged users

Cluster three consisted of 18% of the course participants, which were characterized by the overall highly active participation, typically .5–1 SD above the overall mean (Fig. 2). While these students demonstrate a high volume of activity relating to course assignments and static resources, they also exhibit an active participation in online discussions. Statistical analysis indicated that these students have significantly higher scores on all engagement metrics than students from both clusters one and two (Table 3). We see that they also exhibit the greatest variability in their behavior, indicated by their discriminant function scores (Fig. 3) and standard deviations of trace metrics (Table A.11). The highest variability in behavior is aligned with the previous findings by Kovanović et al. (2015) who also reported the highest variability in behavior for the most active group of students. They are most similar to a broad group of "engaged" students described in a significant number of studies of educational technology use (see Section 2). The diversity in technology use is associated with high level of metacognitive activity (Lust et al., 2013a) and metacognitive monitoring (Hadwin et al., 2007).

Based on the results of survey analysis, we see that cluster three students also show substantial commitment to learning and also awareness of the important aspects that affect learning success. When deciding to enroll into the course, students from cluster three put significantly higher weight on their previous experience with edX platform, as well as the reputation of the given institution and instructors than students from cluster one (Table 6). Moreover, concerning the expectations from the course regarding its relevance to their professional activities, statistical analyses also reveal marginally significant differences between students from cluster three and cluster one, with cluster three students expecting the course to be more relevant to their current work (Table 6). Similarly, survey results reveal that they are significantly more determined to learn than students from cluster one and also more likely to join study groups, and also more often use additional learning materials (Table 8). This indicates the awareness of successful study strategies (Lust et al., 2013a) and the high level of learning regulatory behavior, which are both associated with better learning outcomes (Winne and Hadwin, 1998).

Based on the distribution of their grades, we see that cluster three students were also more likely to obtain the highest grades, in the range of 80%–100% and less likely to have a grade in the 0–10% range. Regarding their motivations for enrolling the course, they are significantly more likely to enroll as a result of their professional occupation and significantly less likely to enroll solely due to general interest in the course topic (Table 7). Thus, it is likely that they adopted mastery goal orientation, and deep approaches to learning (Bliuc et al., 2010), which is associated with higher course success (Trigwell and Prosser, 1991). Finally, regarding their development of cognitive presence, they exhibited higher perceived levels of resolution phase than students from cluster one, similarly to students from cluster two. Among some of the contributing factors to the greater development of cognitive presence is their higher commitment to learning, more focused interest in the course, use of additional course materials, as well as their active participation in online discussions. Given their active use of online discussions, it is likely that they have cohesive conceptions of learning in discussions (Bliuc et al., 2010) and likely perceive them as a valuable learning activity. Given that student-led discussions are shown to better foster the development of cognitive presence than instructor-led discussions (De Wever et al., 2010; Schellens et al., 2007), those students would be good candidates for student moderators to guide discussions into productive directions and assist students from other two clusters (Schrire, 2006; Kovanović et al., 2016b).

6.4. Limitations

There are several limitations in the present study. First of all, as the study is correlational in nature, there are issues related to its internal validity and claims about causality between students' technology use profile and development of three CoI presences, despite CoI instrument being administered after the course. Secondly, while the CoI survey instrument is widely adopted and validated, it is still a self-reported instrument and as such, has limitations which may affect the study findings. While we used the data from a large MOOC with more than 23,000 students, there is a strong self-selection bias of students who decided to complete pre-course and post-course surveys. As it is more likely that committed and engaged students put effort to fill out the questionnaire than disengaged students, it is likely that the observed differences are smaller than the actual ones which limits the power of the conducted statistical tests. We also conducted a large number of statistical tests which increases the chances for Type I errors, which we tried to mitigate this problem by using the statistical correction procedures where possible. However, in the case of statistical analysis of survey responses, this would severely increase the Type II error rate and thus, was not adopted. While the results of our analyses provide a coherent picture of student technology use which is well aligned with the existing literature, there is still a possibility that some of the significant findings are the result of chance alone and not representative of the actual differences between the clusters. Finally, as with the majority of studies of human behavior, there are many potential unaccounted factors affecting student technology use not captured by the learning trace data nor the surveys.

To detect student technology profiles, we adopted clustering procedure which, as an unsupervised machine learning technique, depends on the large number of parameters and modeling decisions (e.g., selected features, pre-processing steps, distance metric, clustering algorithm) which affect the results of the clustering procedure. Moreover, the decisions on the final number of clusters and their interpretation are inherently subjective and may have a significant impact on the study findings. While we based our clustering methodology and the interpretation of identified clusters on the existing literature of student technology use, there is still room for the subjectivity in the interpretation of the clustering results.

7. Conclusions

The results of the present study show several interesting implications for the MOOC research and practice. First of all, our results indicate that simple provision of different technological affordances and resources is not sufficient to secure their successful use by the students. As noted by Clarebout et al. (2013) and Lust et al. (2012, 2013a), the majority of students do not use the available tools appropriately, signaling the lack of metacognitive capacity, skills, and motivation to use the provided technology effectively. This seems to be even more pronounced than in the case of traditional, for-credit online courses where there is strong extrinsic motivation involved and higher uniformity of students and their characteristics.

Based on the results of our analysis, we see that the three different technology use profiles likely require different instructional support. For instance, the first group of students (disengaged users), characterized by the limited use of the available tools and resources, would likely benefit more from instructional interventions focusing on their weak access to online discussions rather addressing their lack of interest in graded course assignments which are out of their scope of participation. Similarly, the second group of students (strategic users), in addition to support for active discussion participation, might benefit from instructional interventions targeting their performance goal orientation and a strong focus on obtaining the certificate. Finally, the third group of students (engaged users), given their strong course commitment and active engagement can be promoted into student moderators, who would then take an instructional role in the discussions, focusing on discussion facilitation and the provision of a more personalized instruction to the less engaged students.

From the research perspective, the present study also shows substantial challenges related to the use of survey data. In particular, the self-selection bias is highly emphasized, which is likely resulting in smaller effect sizes than the ones reported by Kovanović et al. (2015). Similarly, there is likely an additional issue surrounding question interpretation based on the significantly different standards of participation. For example, as cluster one students likely have much lower expectations for discussion engagement than cluster three students, it is very likely that students from these two clusters judge their participation levels according to different baseline standards (e.g., cluster one students might see few discussion posts as appropriate engagement while cluster three students will likely not). While these issues with survey instruments are all well known, the diversity of MOOC participants is making them even more emphasized.

Finally, the present study shows that analysis methods from traditional, for-credit online contexts can be successfully utilized within MOOC contexts. In this study, we adopted an analysis procedure by Kovanović et al. (2015) which was originally developed for the traditional, for-credit online courses. Given the differences in the instructional design and adopted technologies, in the present study, we extracted slightly different engagement metrics from the trace data, which were then used in the same analysis process. With the rising importance of the replicability and adaptation of learning analytics analysis methods across different contexts, the present study provides one particular example of a method which was successfully adopted from traditional for-credit context to the context of MOOCs.

References

Agnihotri, L., Aghababyan, A., Mojarad, S., Riedesel, M., Essa, A., 2015. Mining Login Data for Actionable Student Insight., in: Proceedings of the 8th International Conference on Educational Data Mining (EDM'15), International Educational Data Mining Society. pp. 472–475. URL: http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_233.pdf.

Agresti, A., 2002. Categorical Data Analysis. Second ed., Wiley, New York, NY.

Akyol, Z., Garrison, D.R., 2008. The development of a community of inquiry over time in an online course: Understanding the progression and integration of social, cognitive and teaching presence. Journal of Asynchronous Learning Networks 12, 3–22. URL: http://eric.ed.gov/?id=EJ837483.

Akyol, Z., Garrison, D.R., 2011. Assessing metacognition in an online community of inquiry. The Internet and Higher Education 14, 183–190. doi:10.1016/j.iheduc.2011.01.005.

Anderson, T., Dron, J., 2010. Three generations of distance education pedagogy. The International Review of Research in Open and Distance Learning 12, 80–97. URL: http://www.irrodl.org/index.php/irrodl/article/view/890/.

Anderson, T., Rourke, L., Garrison, D.R., Archer, W., 2001. Assessing Teaching Presence in a Computer Conferencing Context. Journal of Asynchronous Learning Networks 5, 1–17. URL: http://onlinelearningconsortium.org/wp-content/plugins/olc-download/download.php?post_id=4190.

Arbaugh, J., Cleveland-Innes, M., Diaz, S.R., Garrison, D.R., Ice, P., Richardson, J.C., Swan, K.P., 2008. Developing a community of inquiry instrument: Testing a measure of the Community of Inquiry framework using a multi-institutional sample. The Internet and Higher Education 11, 133–136. doi:10.1016/j.iheduc.2008.06.003.

Azevedo, R., 2005. Computer Environments as Metacognitive Tools for Enhancing Learning. Educational Psychologist 40, 193–197. doi:10.1207/s15326985ep4004_1.

Baker, R., Siemens, G., 2013. Educational data mining and learning analytics, in: Sawyer, R.K. (Ed.), The Cambridge handbook of the learning sciences. 2nd edition ed.. Cambridge University Press, Cambridge; New York.

Bender, R., Lange, S., 2001. Adjusting for multiple testing–when and how? Journal of Clinical Epidemiology 54, 343–349.

Bergner, Y., Kerr, D., Pritchard, D.E., 2015. Methodological Challenges in the Analysis of MOOC Data for Exploring the Relationship between Discussion Forum Views and Learning Outcomes, in: Proceedings of the 8th International Conference on Educational Data Mining (EDM'15), pp. 234–241.

Bjork, R.A., Dunlosky, J., Kornell, N., 2013. Self-regulated learning: beliefs, techniques, and illusions. Annual review of psychology 64, 417–444. doi:10.1146/annurev-psych-113011-143823.

Bliuc, A.M., Ellis, R., Goodyear, P., Piggott, L., 2010. Learning through face-to-face and online discussions: Associations between students' conceptions, approaches and academic performance in political science. British Journal of Educational Technology 41, 512–524. doi:10.1111/j.1467-8535.2009.00966.x.

Bock, R.D., 1985. Multivariate Statistical Methods in Behavioral Research. Scientific Software International.

Bovo, A., Sanchez, S., Héguy, O., Duthen, Y., 2013. Analysis of students clustering results based on Moodle log data, in: Proceedings of the 6th International Conference on Educational Data Mining (EDM'13), International Educational Data Mining Society. pp. 306–307.

Bray, J.H., Maxwell, S.E., 1985. Multivariate Analysis of Variance. SAGE.

Butler, D.L., Winne, P.H., 1995. Feedback and Self-Regulated Learning: A Theoretical Synthesis. Review of Educational Research 65, 245–281. doi:10.3102/00346543065003245.

Cerezo, R., Sanchez-Santillan, M., Núñez, J., Paule, M.P., 2015. Different patterns of students interaction with Mooodle and their relationship with achievement, in: Proceedings of the 8th International Conference on Educational Data Mining (EDM'15), International Educational Data Mining Society. pp. 576–577. URL: http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_271.pdf.

Cho, M.H., Kim, B.J., 2013. Students' self-regulation for interaction with others in online learning environments. The Internet and Higher Education 17, 69–75. doi:10.1016/j.iheduc.2012.11.001.

Clarebout, G., Elen, J., Collazo, N.A.J., Lust, G., Jiang, L., 2013. Metacognition and the Use of Tools, in: Azevedo, R., Aleven, V. (Eds.), International Handbook of Metacognition and Learning Technologies. Springer New York. 28, pp. 187–195. doi:10.1007/978-1-4419-5546-3_13.

Cohen, J., 1988. The Analysis of Variance, in: Statistical power analysis for the behavioral sciences. L. Erlbaum Associates, Hillsdale, N.J., pp. 273–406.

Conneely, K.N., Boehnke, M., 2007. So Many Correlated Tests, So Little Time! Rapid Adjustment of P Values for Multiple Correlated Tests. American Journal of Human Genetics 81, 1158–1168.

Coughlan, S., 2015. UK 'biggest online university course'. BBC News URL: http://www.bbc.co.uk/news/education-32721056.

Daniel, J., 2014. Foreword to the Special Section on Massive Open Online Courses MOOCs – Evolution or Revolution? Journal of Online Learning and Teaching 10, i–iv. URL: http://jolt.merlot.org/vol10no1/daniel_foreword_0314.pdf.

Dawson, S., Joksimović, S., Kovanović, V., Gašević, D., Siemens, G., 2015. Recognising learner autonomy: Lessons and reflections from a joint x/c MOOC, in: Proceedings of 2015 HERDSA conference, HERDSA, Melbourne, AU. URL: http://herdsa-2015.p.asnevents.com.au/days/2015-07-08/abstract/22639.

De Wever, B., Keer, H.V., Schellens, T., Valcke, M., 2010. Roles as a structuring tool in online discussion groups: The differential impact of different roles on social knowledge construction. Computers in Human Behavior 26, 516–523. doi:10.1016/j.chb.2009.08.008.

Dennen, V.P., 2008. Pedagogical Lurking: Student Engagement in Non-posting Discussion Behavior. Computers in Human Behavior 24, 1624–1633. doi:10.1016/j.chb.2007.06.003.

Dunlosky, J., Lipko, A.R., 2007. Metacomprehension A Brief History and How to Improve Its Accuracy. Current Directions in Psychological Science 16, 228–232. doi:10.1111/j.1467-8721.2007.00509.x.

Ellis, R.A., Marcus, G., Taylor, R., 2005. Learning through inquiry: student difficulties with online course-based Material. Journal of Computer Assisted Learning 21, 239–252. doi:10.1111/j.1365-2729.2005.00131.x.

Entwistle, N., Tait, H., 1990. Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments. Higher Education 19, 169–194. doi:10.1007/BF00137106.

Entwistle, P.N., 2009. Teaching for Understanding at University: Deep Approaches and Distinctive Ways of Thinking. 1 edition ed., Palgrave Macmillan, Basingstoke, Hampshire ; New York.

Ferguson, R., Clow, D., Beale, R., Cooper, A.J., Morris, N., Bayne, S., Woodgate, A., 2015. Moving Through MOOCS: Pedagogy, Learning Design and Patterns of Engagement, in: Design for Teaching and Learning in a Networked World. Springer International Publishing, pp. 70–84. doi:10.1007/978-3-319-24258-3_6.

Field, A.P., Miles, J., Field, Z., 2012. Discovering statistics using R. SAGE, Los Angeles; London.

Friedman, T.L., 2012. Come the Revolution. The New York Times URL: http://www.nytimes.com/2012/05/16/opinion/friedman-come-the-revolution.html.

Garrison, D.R., 2011. Social presence, in: E-learning in the 21st century: A framework for research and practice. Taylor & Francis, pp. 30–41.

Garrison, D.R., Akyol, Z., 2013. Toward the development of a metacognition construct for communities of inquiry. The Internet and Higher Education 17, 84–89. doi:10.1016/j.iheduc.2012.11.005.

Garrison, D.R., Anderson, T., Archer, W., 1999. Critical inquiry in a text-based environment: Computer conferencing in higher education. The Internet and Higher Education 2, 87–105. doi:10.1016/S1096-7516(00)00016-6.

Garrison, D.R., Anderson, T., Archer, W., 2001. Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education. American Journal of Distance Education 15, 7–23. doi:10.1080/08923640109527071.

Garrison, D.R., Anderson, T., Archer, W., 2010. The first decade of the community of inquiry framework: A retrospective. The Internet and Higher Education 13, 5–9. doi:10.1016/j.iheduc.2009.10.003.

Gašević, D., Adesope, O., Joksimović, S., Kovanović, V., 2015. Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. The Internet and Higher Education 24, 53–65. doi:10.1016/j.iheduc.2014.09.006.

Ge, Y., Dudoit, S., Speed, T.P., 2003. Resampling-based multiple testing for microarray data analysis. Test 12, 1–77. doi:10.1007/BF02595811.

Gelman, A., Hill, J., Yajima, M., 2012. Why We (Usually) Don't Have to Worry About Multiple Comparisons. Journal of Research on Educational Effectiveness 5, 189–211. doi:10.1080/19345747.2011.618213.

Gonyea, R.M., 2005. Self-reported data in institutional research: Review and recommendations. New Directions for Institutional Research 2005, 73–89.

Gordi, T., Khamis, H., 2004. Simple solution to a common statistical problem: Interpreting multiple tests. Clinical Therapeutics 26, 780–786. doi:10.1016/S0149-2918(04)90078-1.

Hadwin, A.F., Nesbit, J.C., Jamieson-Noel, D., Code, J., Winne, P.H., 2007. Examining trace data to explore self-regulated learning. Metacognition and Learning 2, 107–124. doi:10.1007/s11409-007-9016-7.

Hastie, T.J., Tibshirani, R.J., Friedman, J.H., 2013. The elements of statistical learning: data mining, inference, and prediction. Springer, New York, NY.

Hennis, T., Topolovec, S., Poquet, O., de Vries, P., 2016. Who is the Learner: Profiling the Engineering MOOC Student, in: SEFI 44th Annual Conference, Tampere, Finland. URL: http://www.sefi.be/conference-2016/papers/Open_and_Online_Engineering_Education__Engineering_Education_Research/hennis-who-is-the-learner-108_a.pdf.

Hew, K.F., Cheung, W.S., Ng, C.S.L., 2010. Student contribution in asynchronous online discussion: a review of the research and empirical exploration. Instructional Science 38, 571–606. doi:10.1007/s11251-008-9087-0.

Hill, P., 2013. Emerging Student Patterns in MOOCs: A (Revised) Graphical View. URL: http://mfeldstein.com/emerging-student-patterns-in-moocs-a-revised-graphical-view/.

Ho, A.D., Reich, J., Nesterko, S.O., Seaton, D.T., Mullaney, T., Waldo, J., Chuang, I., 2014. HarvardX and MITx: The first year of open online courses, fall 2012-summer 2013. SSRN Scholarly Paper ID 2381263. Social Science Research Network. Rochester, NY. URL: http://papers.ssrn.com/abstract=2381263.

Kizilcec, R.F., Piech, C., Schneider, E., 2013. Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses, in: Proceedings of the Third International Conference on Learning Analytics and Knowledge, ACM, New York, NY, USA. pp. 170–179. doi:10.1145/2460296.2460330.

Kizilcec, R.F., Schneider, E., 2015. Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. ACM Transactions on Computer-Human Interaction 22, 6:1–6:24. doi:10.1145/2699735.

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R., 2016a. Does time-on-task estimation matter? Implications on validity of learning analytics findings. Journal of Learning Analytics 2, 81–110. doi:10.18608/jla.2015.23.6.

Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., Adesope, O., 2015. Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. The Internet and Higher Education 27, 74–89. doi:10.1016/j.iheduc.2015.06.002.

Kovanović, V., Joksimović, S., Gašević, D., Owers, J., Scott, A.M., Woodgate, A., 2016b. Profiling MOOC Course Returners: How Does Student Behavior Change Between Two Course Enrollments?, in: Proceedings of the Third (2016) ACM Conference on Learning @ Scale, ACM, New York, NY, USA. pp. 269–272. doi:10.1145/2876034.2893431.

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., Čukić, I., Vries, P.d., Hatala, M., Dawson, S., Siemens, G., Gašević, D., 2016c. Exploring Communities of Inquiry in Massive Open Online Courses. Manuscript submitted for publication .

Li, N., Kidziński, Ł., Jermann, P., Dillenbourg, P., 2015. MOOC Video Interaction Patterns: What Do They Tell Us?, in: Proceedings of the 10th European Conference on Technology Enhanced Learning, Springer International Publishing. pp. 197–210. doi:10.1007/978-3-319-24258-3_15.

Lust, G., Elen, J., Clarebout, G., 2013a. Regulation of tool-use within a blended course: Student differences and performance effects. Computers & Education 60, 385–395. doi:10.1016/j.compedu.2012.09.001.

Lust, G., Elen, J., Clarebout, G., 2013b. Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. Computers in Human Behavior 29, 2013–2021. doi:10.1016/j.chb.2013.03.014.

Lust, G., Juarez Collazo, N.A., Elen, J., Clarebout, G., 2012. Content management systems: Enriched learning opportunities for all? Computers in Human Behavior 28, 795–808. doi:10.1016/j.chb.2011.12.009.

Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., Clarebout, G., 2011. Tool-use in a blended undergraduate course: In Search of user profiles. Computers & Education 57, 2135–2144. doi:10.1016/j.compedu.2011.05.010.

Milligan, C., Littlejohn, A., Margaryan, A., 2013. Patterns of Engagement in Connectivist MOOCs. Journal of Online Learning and Teaching 9, 149–159.

Nath, R., Pavur, R., 1985. A new statistic in the one-way multivariate analysis of variance. Computational Statistics & Data Analysis 2, 297–315. doi:10.1016/0167-9473(85)90003-9.

Paul, R., 2005. The state of critical thinking today. New Directions for Community Colleges 2005, 27–38. doi:10.1002/cc.193.

Penny, L., Murphy, E., 2009. Rubrics for designing and evaluating online asynchronous discussions. British Journal of Educational Technology 40, 804–820. doi:10.1111/j.1467-8535.2008.00895.x.

Perkins, D.N., 1985. The Fingertip Effect: How Information-Processing Technology Shapes Thinking. Educational Researcher 14, 11–17.

Perneger, T.V., 1998. What's wrong with Bonferroni adjustments. BMJ 316, 1236–1238. doi:10.1136/bmj.316.7139.1236.

Phan, H.P., 2008. Unifying Different Theories of Learning: Theoretical Framework and Empirical Evidence. Educational Psychology 28, 325–340.

Poquet, O., Kovanović, V., Vries, P.d., Hennis, T., Joksimović, S., Gašević, D., Dawson, S., 2016. Social presence in Massive Open Online Courses. Manuscript submitted for publication .

Reich, J., 2015. Rebooting MOOC Research. Science 347, 34–35. doi:10.1126/science.1261627.

Rodrigues, R.L., Ramos, J.L.C., Silva, J.C.S., Gomes, A.S., Souza, F.d.F.d., Maciel, A.M.A., 2016. Discovering Level of Participation in MOOCs through Clusters Analysis, in: 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT), pp. 232–233. doi:10.1109/ICALT.2016.45.

Rothman, K.J., 1990. No adjustments are needed for multiple comparisons. Epidemiology (Cambridge, Mass.) 1, 43–46.

Rourke, L., Anderson, T., Garrison, D.R., Archer, W., 1999. Assessing Social Presence In Asynchronous Text-based Computer Conferencing. The Journal of Distance Education 14, 50–71. URL: http://www.ijede.ca/index.php/jde/article/view/153.

Rovai, A.P., 2007. Facilitating online discussions effectively. The Internet and Higher Education 10, 77–88. doi:10.1016/j.iheduc.2006.10.001.

Rubin, B., Fernandes, R., Avgerinou, M.D., 2013. The effects of technology on the Community of Inquiry and satisfaction with online courses. The Internet and Higher Education 17, 48–57. doi:10.1016/j.iheduc.2012.09.006.

Schellens, T., Keer, H.V., Wever, B.D., Valcke, M., 2007. Scripting by assigning roles: Does it improve knowledge construction in asynchronous discussion groups? International Journal of Computer-Supported Collaborative Learning 2, 225–246. doi:10.1007/s11412-007-9016-2.

Schrire, S., 2006. Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. Computers & Education 46, 49–70. doi:10.1016/j.compedu.2005.04.006.

Senko, C., Hulleman, C.S., Harackiewicz, J.M., 2011. Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. Educational Psychologist 46, 26–47. doi:10.1080/00461520.2011.538646.

Sharma, K., Jermann, P., Dillenbourg, P., 2015. Identifying styles and path towards success in MOOCs, in: Proceedings of the 8th International Conference on Educational Data Mining (EDM'15), Madrid, Spain. pp. 408–411.

Shea, P., Bidjerano, T., 2010. Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. Computers & Education 55, 1721–1731. doi:10.1016/j.compedu.2010.07.017.

Shea, P., Bidjerano, T., 2012. Learning presence as a moderator in the community of inquiry model. Computers & Education 59, 316–326. doi:10.1016/j.compedu.2012.01.011.

Shea, P., Hayes, S., Smith, S.U., Vickers, J., Bidjerano, T., Gozza-Cohen, M., Jian, S.B., Pickett, A., Wilde, J., Tseng, C.H., 2013. Online learner self-regulation: Learning presence viewed through quantitative content- and social network analysis. The International Review of Research in Open and Distance Learning 14, 427–461. URL: http://www.irrodl.org/index.php/irrodl/article/view/1466.

Shea, P., Hayes, S., Smith, S.U., Vickers, J., Bidjerano, T., Pickett, A., Gozza-Cohen, M., Wilde, J., Jian, S., 2012. Learning presence: Additional research on a new conceptual element within the Community of Inquiry (CoI) framework. The Internet and Higher Education 15, 89–95. doi:10.1016/j.iheduc.2011.08.002.

Shea, P., Hayes, S., Uzuner-Smith, S., Gozza-Cohen, M., Vickers, J., Bidjerano, T., 2014. Reconceptualizing the Community of Inquiry framework: Exploratory and confirmatory analysis. The Internet and Higher Education doi:10.1016/j.iheduc.2014.05.002.

Shen, D., Cho, M.H., Tsai, C.L., Marra, R., 2013. Unpacking online learning experiences: Online learning self-efficacy and learning satisfaction. The Internet and Higher Education 19, 10–17. doi:10.1016/j.iheduc.2013.04.001.

Siemens, G., Gasevic, D., 2012. Guest Editorial - Learning and Knowledge Analytics. Educational Technology & Society 15.

Swan, K., Ice, P., 2010. The community of inquiry framework ten years later: Introduction to the special issue. The Internet and Higher Education 13, 1–4. doi:10.1016/j.iheduc.2009.11.003.

Tabachnick, B.G., Fidell, L.S., 2007. Using multivariate statistics. Pearson/Allyn & Bacon, Boston.

Trigwell, K., Prosser, M., 1991. Improving the quality of student learning: The influence of learning context and student approaches to learning on learning outcomes. Higher Education 22, 251–266. doi:10.1007/BF00132290.

Valle, R.d., Duffy, T.M., 2009. Online learning: Learner characteristics and their approaches to managing learning. Instructional Science 37, 129–149. doi:10.1007/s11251-007-9039-0.

Westfall, P.H., Young, S.S., 1993. Resampling-based multiple testing: Examples and methods for p-value adjustment. volume 279. John Wiley & Sons.

Winne, P.H., 1982. Minimizing the black box problem to enhance the validity of theories about instructional effects. Instructional Science 11, 13–28. doi:10.1007/BF00120978.

Winne, P.H., 2006. How software technologies can improve research on learning and bolster school reform. Educational Psychologist 41, 5–17. doi:10.1207/s15326985ep4101_3.

Winne, P.H., Hadwin, A.F., 1998. Studying as self-regulated learning, in: Hacker, D.J., Dunlosky, J., Graesser, A.C. (Eds.), Metacognition in educational theory and practice. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US. The educational psychology series., pp. 277–304.

Winne, P.H., Jamieson-Noel, D., 2002. Exploring students' calibration of self reports about study tactics and achievement. Contemporary Educational Psychology 27, 551–572. doi:10.1016/S0361-476X(02)00006-1.

Wise, A.F., Speer, J., Marbouti, F., Hsiao, Y.T., 2013. Broadening the notion of participation in online discussions: examining patterns in learners' online listening behaviors. Instructional Science 41, 323–343. doi:10.1007/s11251-012-9230-9.

Yen, J.C., Lee, C.Y., 2011. Exploring problem solving patterns and their impact on learning achievement in a blended learning environment. Computers & Education 56, 138–145. doi:10.1016/j.compedu.2010.08.012.

Zhou, M., Winne, P.H., 2012. Modeling academic achievement by self-reported versus traced goal orientation. Learning and Instruction 22, 413–419. doi:10.1016/j.learninstruc.2012.03.004.

Appendix A. Additional materials

In this section, we provide more details about the adopted survey instruments and more detailed results of the performed analyses. First, the relevant parts of pre-course and post-course (Table A.10) surveys are provided in Table A.9 and Table A.10, respectively. We also listed the detailed description of cluster centroids (Table A.11), and mean values of all outcome measures for the three identified clusters (Table A.12). Similarly, the list of standardized discriminant function loadings for the technology use and cognitive presence measures are given in Table A.13 and Table A.14, respectively. Finally, the mean scores for both pre-course and post-course survey questions for the students in the three clusters are shown in Table A.15 and Table A.16, respectively.

Table A.9
Pre-course survey questions used in the present study.

#	Variable	Question text	Type
Basic demographics			
1	`Gender`	What is your gender?	Binary (Male/Female)
2	`Age`	How old are you?	Numeric
Enrollment factors		*Group: How important were the following factors for your choice for this course?*	
3	`ImpFactorPrevEdxExp`	My experience with other courses on edX	5-item Likert-scale
4	`ImpFactorPrevTudelftExp`	My experience with other DelftX courses on edX	5-item Likert-scale
5	`ImpFactorTudelftRep`	The status and reputation of Delft University of Technology	5-item Likert-scale
6	`ImpFactorProfRep`	The status and reputation of the professor(s) involved	5-item Likert-scale
7	`ImpFactorEarningCert`	The possibility of earning a Statement of Accomplishment / Verified Certificate	5-item Likert-scale
8	`ImpFactorCourseUniq`	The uniqueness of this course	5-item Likert-scale
9	`ImpFactorRecommendation`	A recommendation from someone else	5-item Likert-scale
Motivation			
10	`DeterminedToLearn`	To what extent are you determined to learn (more) about functional programming?	5-item Likert-scale
11	`DeterminedToComplete`	To what extent are you determined to complete this course?	5-item Likert-scale
12	`Hours`	How many hours per week do you expect to dedicate to this course?	5-item Likert-scale
13	`CourseInterest`	What describes your interest for registering for this course?	6-item Multiple choice
Background knowledge			
14	`LevelOfEd`	What is the highest level of education that you have?	9-item Multiple choice
15	`EdBackFunctProgRel`	Is your educational background related to (functional) programming?	4-item Multiple choice
16	`PrevExpInTheField`	Do you have professional experience in this field?	4-item Multiple choice
17	`Occupation`	Which of the following best describes your occupation?	8-item Multiple choice
18	`YearsOfWorkExp`	How many years of working experience do you have?	Numeric
Course expectations		*Group: Please rate how important each of the following aspects is for you in this course:*	
19	`ExpectFunChallenge`	The fun and challenge the course offers	5-item Likert-scale
20	`ExpectRelevance`	The practical relevance of the course for me	5-item Likert-scale
21	`ExpectForumFeedbackTeam`	The feedback of the course team on the forum	5-item Likert-scale
22	`ExpectForumPeerInteract`	The interaction with peers on the forum	5-item Likert-scale
Study strategies			
23	`HowLikelyStudyGroup`	How likely do you think it is that you will join a study group?	5-item Likert-scale
24	`HowLikelyMakeFrieds`	How likely do you think it is that you will make friends with others in this course?	5-item Likert-scale
25	`HowLikelyLookExtraMat`	How likely do you think it is that you will look at extra materials for this course?	5-item Likert-scale
26	`HowLikelyPostDisc`	How likely do you think it is that you will post or comment on the discussion board?	5-item Likert-scale
27	`IntentShareExpertise`	How much do you intent to share your expertise with other students in this class?	5-item Likert-scale
28	`PrefAloneVsWithOthers`	Do you prefer to do this course alone or with others?	Binary (Alone/With others)
29	`PrefTeacherVsStudentResp`	To what extent do you prefer a course that is organized by a teacher or institution versus a course where you as a student are responsible for organizing your own learning activities?	Numeric (0–100)
Study support			
30	`SupportBy`	Who supports you in doing this course?	6-item Multiple choice
31	`SupportHow`	How are you supported?	6-item Multiple choice
32	`StudyDuringWorkHrs`	Are you allowed to do this course during work hours?	Binary (Yes/No)
Experience with online environments			
33	`FreqSocialMediaUse`	How often do you use social media (i.e. Facebook, Twitter, Google+, Weibo, etc.)?	5-item Likert-scale
34	`FreqForumUse`	How often do you contribute questions or answers to online forums?	5-item Likert-scale
35	`OnlineClassesTaken`	How many online classes have you ever taken before?	5-item Likert-scale
36	`OnlineClassesComp`	How many online classes have you ever completed?	5-item Likert-scale
Language fluency			
37	`EnglishFluencyLevel`	What is your present level of English fluency?	5-item Likert-scale
38	`EnglishHowComfortable`	How comfortable are you communicating in English?	5-item Likert-scale
39	`EnglishHowOften`	How often do you communicate in English?	5-item Likert-scale

Table A.10
Post-course survey questions used in the present study.

#	Variable	Question text	Type
Course engagement level			
1	`ParticipationLevel`	Since the start of this class, how would you describe your participation level?	6-item Multiple choice
2	`HoursDedicated`	How many hours did you dedicate to this course per week?	Numeric
3	`WorkedHard`	Do you agree: “I worked really hard at things I didn’t understand.”	5-item Likert-scale
4	`FamilyWorkObligationsAside`	Do you agree: “I had to put aside family/work obligations because of the course.”	5-item Likert-scale
5	`PlannedAndOrganizedLearning`	Do you agree: “I planned and organized my learning well.”	5-item Likert-scale
6	`LookedAtExtraMaterials`	How often have you looked at extra materials for this course?	5-item Likert-scale
Social interactions			
7	`ForumHowOften`	How often did you take a look and read discussions on the forum?	5-item Likert-scale
8	`PostedInDiscussions`	Howe often have you posted a comment or question on the course discussion board?	5-item Likert-scale
9	`SharedExpertiseWithOthers`	Do you agree: “I shared my expertise with other students.”	5-item Likert-scale
10	`IntendedToShareMore`	Do you agree: “I intended to share my expertise with other students more.”	5-item Likert-scale
11	`ConnWithCourseTeam`	How often have you contacted the instructors?	5-item Likert-scale
12	`ConnWithOtherStudents`	Do you agree: “I connected with other students.”	5-item Likert-scale
13	`ConnWithNewFriends`	How often have you connected with new friends from this course?	5-item Likert-scale
14	`WouldLikedConnectMore`	Do you agree: “I would have liked to connect with other students more.”	5-item Likert-scale
15	`PartInStudyGroup`	How often have you participated a study group?	5-item Likert-scale
Challenges to course participation		*Group: Did any of the following negatively affect your participation in the course?*	
16	`PersonalMedicalIssue`	A personal or medical issue	5-item Likert-scale
17	`ProfessionalObligation`	Important professional obligations	5-item Likert-scale
18	`NotGettingFeedback`	Not getting (the right) feedback when I needed it	5-item Likert-scale
19	`FeelingLonely`	The feeling of loneliness or missing interaction with peers	5-item Likert-scale
20	`TechOrAccessibilityIssue`	Technical or accessibility issues	5-item Likert-scale
21	`KeepingUpWithCoursePace`	Problems with having to keep up with the pace of the course	5-item Likert-scale
22	`FamilyObligation`	Important family obligations	5-item Likert-scale
23	`AnotherCourseObligation`	Obligations from another (online) course or study	5-item Likert-scale
Technical challenges to course participation		*Group: how much did each of the technical issues affect your participation?*	
24	`SlowInternet`	Slow Internet	5-item Likert-scale
25	`InternetNetworkProblems`	Internet network problems	5-item Likert-scale
26	`ElectricityNetworkProblems`	Electricity network problems	5-item Likert-scale
27	`NotHavingAccessPers.Comp.`	Not having access to a personal computer	5-item Likert-scale
28	`HardwareProblems`	Hardware problems	5-item Likert-scale
29	`NotHavingMobileAccess`	Not being able to access it on my mobile device	5-item Likert-scale
30	`AccessibilityUsability`	Usability or accessibility of the website (i.e. because of a disability)	5-item Likert-scale
Course appropriate		*Group: How much do you agree with the following statements?*	
31	`CourseExpectationsRealistic`	“My expectations about the course were realistic”	5-item Likert-scale
32	`CourseMeetYourExpectations`	Did the course meet your expectations?	3-item Likert-scale
33	`LevelOfEnglishAppropriate`	“My level of English was appropriate for this course.”	5-item Likert-scale
34	`CourseRelevantToOccupation`	“The course was relevant for my profession or current occupation.”	5-item Likert-scale
35	`PriorKnowledgeMadeItEasier`	“My prior knowledge made it easier to complete/understand assignments/lectures.”	5-item Likert-scale
Course support			
36	`CourseInspiredToStudy`	Do you agree: “This MOOC inspired me to continue studying in this field.”	5-item Likert-scale
37	`ForumWasHelpfulForMe`	Do you agree: “The course forum was helpful for me.”	5-item Likert-scale
38	`OthersHelpedMeInTheCourse`	Do you agree: “Others helped me in this course.”	5-item Likert-scale
39	`ReceivedSupportStudents`	Do you agree: “I received support from other students.”	5-item Likert-scale
40	`ReceivedSupportCourseTeam`	Do you agree: “I received support from the Teaching Assistants or teacher(s).”	5-item Likert-scale
41	`SupportFromCourseTeamForums`	How would you rate the support from the course team on the forum?	5-item Likert-scale
42	`OthersCouldHelpMore`	Do you agree: “I think others could have helped me more in this course.”	5-item Likert-scale
43	`ForumCouldBeMoreHelpful`	Do you agree: “I think the course forum could have been more helpful for me.”	5-item Likert-scale
Course design and quality evaluation			
44	`DifficultyLevelOfTheCourse`	The difficulty level of the course was:	5-item Likert-scale
45	`AmountOfWorkRequired`	The amount of work required for the course was:	5-item Likert-scale
46	`PaceOfTheCourse`	The pace of the course was:	5-item Likert-scale
47	`DurationOfTheCourse`	The duration of the course (number of weeks) was:	5-item Likert-scale
48	`LecturesExercisesBalance`	How would you rate the balance between lectures and exercises?	5-item Likert-scale
49	`CourseOverallQuality`	How would you rate the overall quality of the course?	5-item Likert-scale
50	`AssignmentAndExamQuality`	How would you rate the quality of the assignments and exams?	5-item Likert-scale
51	`VideoLecturesQuality`	How would you rate the quality of the video lectures?	5-item Likert-scale
52	`FeedbackQuality`	How would you rate the feedback on completed quizzes and assignments?	5-item Likert-scale
53	`EdxEaseOfUseAndQuality`	How would you rate The ease of use and overall quality of the edX platform?	5-item Likert-scale

Table A.11
The centroids of the three identified clusters.

#	Variable	Cluster 1 (N = 15,868)	Cluster 2 (N = 3,532)	Cluster 3 (N = 4,248)
	Course access			
1	`OpenHome`	2.55 (2.37)	7.91 (5.78)	27.94 (112.26)
2	`OpenWiki`	0.28 (0.55)	1.20 (1.58)	2.79 (3.63)
3	`OpenProgress`	0.55 (1.34)	3.47 (6.01)	26.76 (53.78)
4	`OpenSyllabus`	0.49 (1.12)	1.41 (1.76)	4.67 (33.84)
5	`OpenCourseware`	1.92 (2.12)	7.40 (5.82)	31.54 (44.43)
6	`OpenCoursewareItem`	5.38 (12.92)	27.16 (18.46)	113.35 (351.48)
	Assignments			
7	`AssignmentStart`	0.60 (2.10)	7.83 (11.18)	38.66 (41.44)
8	`AssignmentSubmit`	1.13 (4.68)	20.72 (30.04)	179.20 (109.74)
	Video lectures			
9	`VideoLoad`	3.74 (4.83)	21.66 (14.18)	60.99 (37.27)
10	`VideoPlay`	8.26 (27.82)	44.85 (67.97)	215.82 (1010.97)
11	`VideoPause`	3.74 (9.09)	29.12 (114.00)	90.70 (175.11)
12	`VideoChangeSpeed`	0.29 (0.97)	2.02 (4.62)	6.11 (12.64)
13	`VideoShowSubs`	2.76 (4.01)	15.88 (13.06)	41.64 (36.23)
	Course navigation			
14	`ModuleNext`	1.74 (2.45)	10.33 (7.63)	30.95 (24.74)
15	`ModulePrev`	0.09 (0.34)	1.05 (1.57)	3.54 (6.35)
16	`ModuleJump`	2.55 (4.84)	17.10 (16.25)	62.34 (51.50)
	Discussion access			
17	`ThreadAccess`	0.49 (2.94)	3.80 (9.53)	70.97 (307.41)
18	`ThreadAccessInline`	0.05 (0.26)	0.76 (1.33)	6.61 (9.97)
19	`DiscussionSearch`	0.03 (0.33)	0.12 (0.54)	3.71 (11.72)
	Discussion contribution			
20	`DiscussionsStarted`	0.00 (0.00)	0.00 (0.02)	0.16 (0.90)
21	`QuestionsStarted`	0.00 (0.00)	0.00 (0.00)	0.14 (0.75)
22	`CommentsWritten`	0.00 (0.03)	0.00 (0.07)	1.62 (19.60)
23	`ThreadsCharsAvg`	0.00 (0.00)	0.03 (2.02)	61.57 (215.54)
24	`CommentsCharsAvg`	0.02 (0.81)	0.11 (2.08)	55.17 (154.86)
25	`UpvotesGiven`	0.00 (0.06)	0.03 (0.28)	1.10 (6.61)
26	`CommentsEnd.Given`	0.00 (0.00)	0.00 (0.00)	0.18 (7.81)
	Discussion reputation			
27	`CommentsReceived`	0.00 (0.00)	0.00 (0.02)	1.62 (8.50)
28	`UpvotesReceived`	0.00 (0.00)	0.00 (0.00)	0.86 (12.10)
29	`CommentsEnd.Rec.`	0.00 (0.00)	0.00 (0.00)	0.15 (3.44)

Table A.12
Scores on the outcome variables of the identified clusters.

#	Variable	Cluster 1	Cluster 2	Cluster 3
	Course grades			
1	`FinalGrade`	0.09 (0.23)	0.10 (0.24)	0.11 (0.26)
	CoI: Perceived levels of cognitive presence			
1	`TriggeringEventLevel`	3.8 (0.81)	3.9 (0.72)	4.0 (0.73)
2	`ExplorationLevel`	3.6 (0.67)	3.6 (0.66)	3.7 (0.66)
3	`IntegrationLevel`	3.7 (0.74)	3.7 (0.67)	3.7 (0.70)
4	`ResolutionLevel`	3.5 (0.84)	3.7 (0.85)	3.7 (0.79)
	CoI: Perceived levels of teaching presence			
5	`Org.DesignLevel`	3.9 (0.77)	3.9 (0.78)	4.0 (0.73)
6	`FacilitationLevel`	3.7 (0.77)	3.7 (0.75)	3.8 (0.73)
7	`DirectInst.Level`	3.4 (0.83)	3.4 (0.84)	3.5 (0.78)
	CoI: Perceived levels of social presence			
8	`AffectiveExp.Level`	2.9 (0.69)	3.0 (0.73)	3.0 (0.67)
9	`OpenComm.Level`	3.2 (0.78)	3.3 (0.78)	3.2 (0.81)
10	`GroupCohesionLevel`	3.0 (0.60)	3.2 (0.59)	3.1 (0.60)

Table A.13
Standardized DFA of clustering variables.

#	Variable	LD1	LD2
	Course access		
1	`OpenHome`	0.21	0.31
2	`OpenWiki`	0.12	0.40
3	`OpenProgress`	-0.16	-0.26
4	`OpenSyllabus`	-0.32	-0.33
5	`OpenCourseware`	-0.04	-0.22
6	`OpenCoursewareItem`	-0.02	-0.14
	Assignments		
7	`AssignmentStart`	0.30	-0.01
8	`AssignmentSubmit`	0.45	-0.83
	Video lectures		
9	`VideoLoad`	0.24	0.18
10	`VideoPlay`	0.01	-0.12
11	`VideoPause`	0.02	0.05
12	`VideoChangeSpeed`	0.09	0.05
13	`VideoShowSubs`	0.08	0.13
	Course navigation		
14	`ModuleNext`	0.43	0.73
15	`ModulePrev`	-0.01	-0.27
16	`ModuleJump`	0.27	0.55
	Discussion access		
17	`ThreadAccess`	-0.08	-0.10
18	`ThreadAccessInline`	-0.17	-0.29
19	`DiscussionSearch`	-0.08	-0.11
	Discussion contribution		
20	`DiscussionsStarted`	-0.00	0.01
21	`QuestionsStarted`	0.05	-0.06
22	`CommentsWritten`	0.19	0.07
23	`ThreadsCharsAvg`	0.10	-0.09
24	`CommentsCharsAvg`	0.15	-0.14
25	`UpvotesGiven`	-0.04	-0.04
26	`CommentsEnd.Given`	0.03	-0.12
	Discussion reputation		
27	`CommentsReceived`	-0.06	-0.01
28	`UpvotesReceived`	-0.05	0.45
29	`CommentsEnd.Rec.`	-0.09	-0.38
	Variance Explained	0.97	0.03

Table A.14
Standardized DFA loadings of cognitive presence variables.

#	Variable	LD1	LD2
1	`TriggeringEventLevel`	-0.32	-0.64
2	`ExplorationLevel`	-0.14	-0.79
3	`IntegrationLevel`	0.81	0.10
4	`ResolutionLevel`	-1.12	0.54
	Variance Explained	0.94	0.06

Table A.15
The average scores of answers to the course surveys of students in the three identified clusters.

		Cluster 1 (Students = 15,868)		Cluster 2 (Students = 3,532)		Cluster 3 (Students = 4,248)	
#	Variable	Mean (SD)	Responses (Cluster %)	Mean (SD)	Responses (Cluster %)	Mean (SD)	Responses (Cluster %)
Basic demographics							
1	`Gender`	1.06 (0.26)	2,505 (16%)	1.07 (0.30)	619 (18%)	1.06 (0.28)	741 (17%)
2	`Age`	33.46 (10.56)	2,449 (15%)	32.49 (9.96)	600 (17%)	33.76 (10.79)	724 (17%)
Enrollment factors							
3	`ImpFactorPrevEdxExp`	2.21 (1.32)	2,606 (16%)	2.12 (1.34)	648 (18%)	1.98 (1.27)	773 (18%)
4	`ImpFactorPrevTudelftExp`	1.40 (0.83)	2,592 (16%)	1.33 (0.77)	646 (18%)	1.35 (0.82)	770 (18%)
5	`ImpFactorTudelftRep`	1.85 (1.11)	2,597 (16%)	1.78 (1.10)	647 (18%)	1.74 (1.08)	770 (18%)
6	`ImpFactorProfRep`	2.75 (1.46)	2,598 (16%)	2.84 (1.50)	648 (18%)	2.91 (1.47)	774 (18%)
7	`ImpFactorEarningCert`	1.96 (1.18)	2,605 (16%)	1.97 (1.19)	648 (18%)	1.95 (1.18)	772 (18%)
8	`ImpFactorCourseUniq`	3.09 (1.21)	2,604 (16%)	3.14 (1.23)	649 (18%)	3.09 (1.23)	774 (18%)
9	`ImpFactorRecommendation`	1.88 (1.26)	2,592 (16%)	1.90 (1.24)	645 (18%)	1.95 (1.27)	772 (18%)
Motivation							
10	`DeterminedToLearn`	3.92 (0.82)	2,778 (18%)	4.00 (0.81)	685 (19%)	4.02 (0.79)	842 (20%)
11	`DeterminedToComplete`	3.73 (0.83)	2,777 (18%)	3.77 (0.85)	685 (19%)	3.77 (0.83)	843 (20%)
12	`Hours`	2.29 (0.78)	2,778 (18%)	2.35 (0.80)	683 (19%)	2.36 (0.81)	842 (20%)
13	`CourseInterest`	4.61 (1.96)	2,745 (17%)	4.39 (2.03)	678 (19%)	4.39 (2.02)	823 (19%)
Background knowledge							
14	`LevelOfEd`	3.02 (1.38)	2,495 (17%)	3.08 (1.42)	615 (17%)	2.93 (1.24)	745 (18%)
15	`EdBackFunctProgRel`	1.84 (1.00)	2,498 (16%)	1.86 (0.99)	615 (17%)	1.86 (1.00)	742 (17%)
16	`PrevExpInTheField`	1.94 (1.06)	2,496 (16%)	2.00 (1.05)	615 (17%)	2.07 (1.07)	743 (17%)
17	`Occupation`	3.03 (1.33)	2,496 (16%)	2.94 (1.23)	616 (17%)	2.99 (1.18)	743 (17%)
18	`YearsOfWorkExp`	11.85 (12.99)	1,726 (11%)	10.97 (8.30)	439 (12%)	12.01 (9.47)	538 (13%)
Course expectations							
19	`ExpectFunChallenge`	3.90 (0.83)	2,772 (17%)	3.95 (0.81)	685 (19%)	3.88 (0.85)	843 (20%)
20	`ExpectRelevance`	3.72 (0.95)	2,772 (17%)	3.80 (0.92)	685 (19%)	3.81 (0.92)	843 (20%)
21	`ExpectForumFeedbackTeam`	2.58 (1.01)	2,768 (17%)	2.62 (1.04)	685 (19%)	2.62 (1.05)	840 (20%)
22	`ExpectForumPeerInteract`	2.29 (1.00)	2,769 (17%)	2.31 (1.04)	685 (19%)	2.30 (1.01)	841 (20%)
Study strategies							
23	`HowLikelyStudyGroup`	1.80 (0.94)	2,770 (17%)	1.87 (0.98)	685 (19%)	1.90 (0.98)	842 (20%)
24	`HowLikelyMakeFrieds`	1.86 (0.89)	2,770 (17%)	1.86 (0.91)	685 (19%)	1.90 (0.91)	841 (20%)
25	`HowLikelyLookExtraMat`	3.61 (0.91)	2,771 (17%)	3.65 (0.90)	685 (19%)	3.63 (0.91)	843 (20%)
26	`HowLikelyPostDisc`	2.73 (1.00)	2,770 (17%)	2.79 (1.01)	685 (19%)	2.80 (0.99)	841 (20%)
27	`IntentShareExpertise`	3.04 (0.86)	2,569 (16%)	3.02 (0.91)	638 (18%)	3.08 (0.87)	765 (18%)
28	`PrefAloneVsWithOthers`	1.23 (0.42)	2,574 (16%)	1.26 (0.44)	637 (18%)	1.25 (0.43)	763 (18%)
29	`PrefTeacherVsStudentResp`	42.40 (24.26)	2,599 (16%)	40.68 (24.80)	647 (18%)	43.33 (23.42)	774 (18%)
Study support							
30	`SupportBy`	1.19 (0.75)	2,479 (16%)	1.21 (0.85)	616 (17%)	1.18 (0.74)	739 (17%)
31	`SupportHow`	3.44 (1.68)	186 (1%)	3.65 (1.57)	43 (1%)	3.48 (1.72)	52 (1%)
32	`StudyDuringWorkHrs`	1.72 (0.45)	1,748 (11%)	1.71 (0.45)	444 (13%)	1.70 (0.46)	544 (13%)
Experience with online environments							
33	`FreqSocialMediaUse`	3.03 (1.19)	2,575 (16%)	3.03 (1.16)	639 (18%)	3.09 (1.25)	768 (18%)
34	`FreqForumUse`	2.51 (0.82)	2,575 (16%)	2.55 (0.83)	639 (18%)	2.50 (0.78)	768 (18%)
35	`OnlineClassesTaken`	3.36 (1.20)	2,574 (16%)	3.29 (1.28)	639 (18%)	3.26 (1.29)	768 (18%)
36	`OnlineClassesComp`	3.48 (1.48)	2,262 (14%)	3.41 (1.49)	538 (15%)	3.36 (1.47)	645 (15%)
Language fluency							
37	`EnglishFluencyLevel`	4.14 (0.98)	2,483 (16%)	4.18 (0.97)	616 (17%)	4.19 (0.94)	740 (17%)
38	`EnglishHowComfortable`	4.34 (0.90)	2,483 (16%)	4.41 (0.83)	616 (17%)	4.38 (0.86)	740 (17%)
39	`EnglishHowOften`	3.87 (1.03)	2,483 (16%)	3.96 (0.99)	616 (17%)	3.91 (1.00)	740 (17%)

Table A.16
The average scores of answers to the course surveys of students in the three identified clusters.

		Cluster 1 (Students = 15,868)		Cluster 2 (Students = 3,532)		Cluster 3 (Students = 4,248)	
#	Variable	Mean (SD)	Responses (Cluster %)	Mean (SD)	Responses (Cluster %)	Mean (SD)	Responses (Cluster %)
Course engagement level							
1	`ParticipationLevel`	5.16 (0.91)	615 (4%)	5.30 (0.85)	164 (5%)	5.27 (0.79)	208 (5%)
2	`HoursDedicated`	6.27 (4.10)	532 (3%)	6.45 (6.89)	139 (4%)	6.50 (4.05)	188 (4%)
3	`WorkedHard`	1.15 (0.50)	556 (4%)	1.18 (0.47)	154 (4%)	1.17 (0.53)	196 (5%)
4	`FamilyWorkObligationsAside`	1.09 (0.36)	556 (4%)	1.06 (0.37)	154 (4%)	1.11 (0.46)	196 (5%)
5	`PlannedAndOrganizedLearning`	1.23 (0.56)	557 (4%)	1.17 (0.51)	154 (4%)	1.25 (0.53)	196 (5%)
6	`LookedAtExtraMaterials`	3.25 (1.05)	557 (4%)	3.17 (1.00)	155 (4%)	3.46 (0.90)	196 (5%)
Social interactions							
7	`ForumHowOften`	1.54 (0.81)	557 (4%)	1.46 (0.73)	154 (4%)	1.56 (0.81)	196 (5%)
8	`PostedInDiscussions`	3.27 (0.90)	561 (4%)	3.28 (0.86)	152 (4%)	3.36 (0.87)	198 (5%)
9	`SharedExpertiseWithOthers`	3.44 (0.97)	570 (4%)	3.48 (0.99)	157 (4%)	3.42 (1.01)	201 (5%)
10	`IntendedToShareMore`	2.29 (1.03)	570 (4%)	2.31 (0.96)	156 (4%)	2.35 (1.01)	201 (5%)
11	`ConnWithCourseTeam`	3.04 (0.94)	570 (4%)	3.01 (0.96)	155 (4%)	3.05 (0.90)	201 (5%)
12	`ConnWithOtherStudents`	1.74 (0.99)	559 (4%)	1.71 (1.03)	154 (4%)	1.83 (0.95)	196 (5%)
13	`ConnWithNewFriends`	1.80 (1.04)	559 (4%)	1.79 (1.09)	154 (4%)	2.03 (1.06)	196 (5%)
14	`WouldLikedConnectMore`	2.59 (1.30)	559 (4%)	2.50 (1.37)	154 (4%)	2.72 (1.29)	196 (5%)
15	`PartInStudyGroup`	3.45 (1.16)	560 (4%)	3.38 (1.29)	155 (4%)	3.42 (1.17)	195 (5%)
Challenges to course participation							
16	`PersonalMedicalIssue`	1.29 (0.72)	594 (4%)	1.31 (0.78)	160 (5%)	1.21 (0.53)	203 (5%)
17	`ProfessionalObligation`	2.33 (1.18)	595 (4%)	2.31 (1.14)	160 (5%)	2.33 (1.11)	204 (5%)
18	`NotGettingFeedback`	1.44 (0.81)	594 (4%)	1.42 (0.87)	161 (5%)	1.42 (0.82)	203 (5%)
19	`FeelingLonely`	1.29 (0.65)	594 (4%)	1.29 (0.60)	160 (5%)	1.32 (0.67)	203 (5%)
20	`TechOrAccessibilityIssue`	1.15 (0.50)	594 (4%)	1.20 (0.59)	161 (5%)	1.15 (0.55)	203 (5%)
21	`KeepingUpWithCoursePace`	1.98 (1.07)	594 (4%)	1.86 (1.06)	160 (5%)	1.84 (1.03)	203 (5%)
22	`FamilyObligation`	2.01 (1.10)	595 (4%)	1.96 (1.10)	160 (5%)	1.99 (1.01)	204 (5%)
23	`AnotherCourseObligation`	1.64 (0.97)	596 (4%)	1.64 (1.00)	160 (5%)	1.60 (1.02)	203 (5%)
Technical challenges to course participation							
24	`SlowInternet`	1.18 (0.56)	588 (4%)	1.18 (0.61)	157 (4%)	1.17 (0.48)	202 (5%)
25	`InternetNetworkProblems`	1.19 (0.56)	588 (4%)	1.24 (0.70)	157 (4%)	1.20 (0.63)	202 (5%)
26	`ElectricityNetworkProblems`	1.05 (0.33)	588 (4%)	1.06 (0.34)	157 (4%)	1.00 (0.07)	202 (5%)
27	`NotHavingAccessPers.Comp.`	1.07 (0.37)	588 (4%)	1.10 (0.49)	157 (4%)	1.03 (0.32)	202 (5%)
28	`HardwareProblems`	1.05 (0.32)	587 (4%)	1.01 (0.08)	157 (4%)	1.05 (0.36)	202 (5%)
29	`NotHavingMobileAccess`	1.24 (0.64)	589 (4%)	1.33 (0.78)	157 (4%)	1.21 (0.60)	202 (5%)
30	`AccessibilityUsability`	1.17 (0.55)	588 (4%)	1.11 (0.45)	157 (4%)	1.24 (0.63)	202 (5%)
Course appropriate							
31	`CourseExpectationsRealistic`	3.48 (0.88)	572 (4%)	3.54 (0.83)	157 (4%)	3.61 (0.84)	201 (5%)
32	`CourseMeetYourExpectations`	4.57 (0.66)	573 (4%)	4.53 (0.72)	158 (4%)	4.50 (0.69)	201 (5%)
33	`LevelOfEnglishAppropriate`	3.30 (1.11)	572 (4%)	3.44 (1.03)	158 (4%)	3.25 (1.10)	201 (5%)
34	`CourseRelevantToOccupation`	3.55 (1.01)	573 (4%)	3.65 (0.92)	158 (4%)	3.61 (0.95)	201 (5%)
35	`PriorKnowledgeMadeItEasier`	1.95 (0.72)	502 (3%)	1.83 (0.70)	138 (4%)	1.96 (0.70)	176 (4%)
Course support							
36	`CourseInspiredToStudy`	3.69 (0.74)	545 (3%)	3.69 (0.72)	150 (4%)	3.60 (0.78)	191 (4%)
37	`ForumWasHelpfulForMe`	3.24 (0.61)	550 (3%)	3.17 (0.54)	155 (4%)	3.14 (0.55)	194 (5%)
38	`OthersHelpedMeInTheCourse`	3.28 (0.61)	551 (3%)	3.29 (0.57)	154 (4%)	3.28 (0.60)	194 (5%)
39	`ReceivedSupportStudents`	2.87 (0.53)	553 (3%)	2.87 (0.42)	154 (4%)	2.84 (0.47)	193 (5%)
40	`ReceivedSupportCourseTeam`	4.01 (0.84)	553 (3%)	4.07 (0.83)	154 (4%)	4.04 (0.81)	193 (5%)
41	`SupportFromCourseTeamForums`	3.74 (1.05)	552 (3%)	3.69 (1.10)	154 (4%)	3.81 (0.99)	193 (5%)
42	`OthersCouldHelpMore`	3.54 (1.07)	550 (3%)	3.64 (1.02)	154 (4%)	3.68 (0.92)	193 (5%)
43	`ForumCouldBeMoreHelpful`	4.00 (0.92)	552 (3%)	4.15 (0.81)	154 (4%)	4.09 (0.83)	192 (5%)
Course design and quality evaluation							
44	`DifficultyLevelOfTheCourse`	3.10 (1.10)	543 (3%)	3.14 (1.03)	153 (4%)	3.28 (1.04)	187 (4%)
45	`AmountOfWorkRequired`	3.69 (0.95)	535 (3%)	3.74 (0.89)	149 (4%)	3.68 (0.90)	181 (4%)
46	`PaceOfTheCourse`	3.82 (0.97)	553 (3%)	3.72 (1.00)	154 (4%)	3.70 (0.98)	193 (5%)
47	`DurationOfTheCourse`	3.95 (1.07)	552 (3%)	4.05 (1.07)	154 (4%)	4.09 (1.00)	192 (5%)
48	`LecturesExercisesBalance`	2.50 (1.02)	557 (4%)	2.53 (1.06)	154 (4%)	2.69 (0.97)	196 (5%)
49	`CourseOverallQuality`	2.68 (1.04)	558 (4%)	2.66 (1.07)	154 (4%)	2.87 (1.06)	196 (5%)
50	`AssignmentAndExamQuality`	1.94 (0.97)	566 (4%)	1.90 (1.02)	154 (4%)	2.06 (1.08)	200 (5%)
51	`VideoLecturesQuality`	1.86 (1.06)	566 (4%)	1.75 (0.99)	154 (4%)	1.92 (1.05)	199 (5%)
52	`FeedbackQuality`	2.76 (1.06)	558 (4%)	2.90 (1.07)	154 (4%)	2.96 (1.03)	196 (5%)
53	`EdxEaseOfUseAndQuality`	2.62 (1.02)	558 (4%)	2.73 (1.04)	154 (4%)	2.93 (0.92)	196 (5%)

7.3 Summary

In this chapter, we examined the use of learning analytics for assessing students' learning experience – in particular, the development of cognitive presence – using the trace data of students' interactions with the learning platform. Building on the work presented in Chapter four, we developed a novel learning analytics system for assessing students' cognitive presence within MOOCs by examining their learning self-regulation. After the validation of the CoI model within the MOOC context (Chapter six), we identified key learning strategies based on students' use of available tools and resources, which are indicative of their prior knowledge and beliefs (i.e., knowledge of effective learning strategies, self-efficacy beliefs), learning strategies and self-regulatory processes (i.e., metacognitive monitoring and control). Given the importance of self-regulation and metacognition on student inquiry-based learning (Akyol & Garrison, 2011a; Garrison & Akyol, 2013; Shea & Bidjerano, 2010, 2012; Shea et al., 2012), the model provides the view on the student cognitive presence from the personal, reflective dimension of inquiry-based learning (Garrison et al., 2001).

Similar to analytics models described in Chapter three and Chapter four, the analytics system, described in this chapter, represents an actual analytics implementation on the top, assessment approach layer of the cognitive presence assessment model (Chapter two). The system itself was developed by adjusting the assessment approach described in Chapter four, so that the new model can reflect differences in course pedagogy and adopted learning platform. This includes the reevaluation of the task model to include the list of activities specified by the instructional design of the target MOOC course. After the learning activities have been specified, we designed a list of relevant metrics of student learning, drawing on the published literature in the MOOC field and on the capabilities of the MOOC platform used in the study. For instance, given the challenges of time-on-task estimation covered in Chapter five and also the limitations of the target MOOC platform regarding the collection of discussion-related trace data, we decided to exclude time-on-task measures from the final analytics model implementation. To compensate for the exclusion of time-on-task measures, we focused instead on the measures relating to the quality of the discussion contributions (e.g., average number of characters per new discussion/reply).

An important contribution of the present chapter – aside from examining the cognitive presence in relation to students' technology use – is the identification of challenges related to the use of self-reported measures for providing a baseline assessment of students' levels of cognitive presence. As a result, the work presented in this chapter also provides further evidence of the importance of adopting more objective measures of cognitive presence and other latent constructs than it can be achieved with self-report instruments. Therefore, our future work in this domain would be focused on the use of the classification system developed in Chapter three to provide a baseline assessment of students' cognitive presence using analysis of discussion transcripts. However, to achieve this, we first need to examine the accuracy of the classification system within the MOOC context, given the substantial differences regarding with regards to the discussion facilitation and direct instruction.

8 Conclusions and future directions

> No book can ever be finished. While working on it we learn just enough to find it immature the moment we turn away from it.
>
> — Karl R. Popper, *The Open Society and Its Enemies*

The overarching idea of this thesis is to use data collected by learning environments to provide assessment and a better understanding of students' cognitive presence using techniques from the field of learning analytics. In this thesis, we presented several analytics systems that provide valuable insights into the development of cognitive presence from several viewpoints, acknowledging that cognitive presence is a multifaceted construct that captures both personal and social aspects of learning.

In this chapter, we briefly summarize the main findings and contributions of the work presented in this thesis according to the key research goals and questions stated in Section 1.1. Given the practical nature of learning analytics research, we dedicate special attention to the impact of the present work and its implications, both for research and practice of online education. We also examine the potentially fruitful avenues for future work. Finally, we conclude the thesis with some concluding remarks.

8.1 Impact of the present work

8.1.1 RQ 1: Development of cognitive presence assessment model

In Chapter two, we described the conceptual model for cognitive presence assessment. Besides serving as a foundation for the present thesis, the primary motivation behind the model development were 1) to provide an overview of the core elements of any analytics system for cognitive presence assessment and 2) to serve as a template for the specification of any system for analytics of cognitive presence.

There are several benefits of explicit specification of the components of the analytics model. First of all, the educational theory layer of the assessment model provides a clear connection to the CoI model which serves as a theoretical foundation for any analytics-based assessment system.

Secondly, the educational technology layer of the model precisely defines the types of data sources which are required for the development of cognitive presence analytics system. This is important as often the implementation of learning analytics systems involves the coordination of several academic and administrative departments and the explicit requirements regarding the input data eases the collaboration around data access and preprocessing. Thirdly, and most importantly, the assessment framework layer of the model explicitly frames the development of cognitive presence assessment in the form of three core elements drawn from the evidence-centered design framework (i.e., student, task, and evidence models). As such, any actual implementation on the final, the analytics approach layer can be framed in terms of those three components which eases adoption and adjustment of the model for different learning contexts.

Secondly, by being based on the widely-used ECD framework, the cognitive presence assessment model is theoretically grounded in the existing educational assessment literature. In particular, the use of student, task, and evidence models provides strong methodological foundation for development of new learning analytics assessment systems and also for their adjustment for the new learning settings. In this regard, the present thesis serves as one of the first examples on how the development of learning analytics can be linked with the existing literature on educational assessment – such as the ECD framework – and also with different theories of teaching and learning – such as the CoI model.

8.1.2 RQ 2: Data-informed operationalization of cognitive presence

The next significant contribution of the present thesis is the data-informed operationalization of the different cognitive presence phases presented in Chapter three. As explained by Rourke et al. (2001), the results of the content analysis depend on the complexity of a coding scheme, as well on the quality and completeness of the provided indicators of the coding scheme. Given that the original coding scheme by Garrison et al. (2001) was targeted towards trained educational researchers, it is reasonably succinct and abstract. For instance, the triggering event phase was characterized as "evocative", and identified using two indicators: 1) "recognizing the problem", described as *"presenting background information that culminates in a question"* (Garrison et al., 2001, p. 15); and 2) "sense of puzzlement", described as *"asking questions"* or *"messages that take discussion in new direction."* (Garrison et al., 2001, p. 15) Although this specification of the triggering event phase is certainly reasonable, it also leaves much space for personal interpretation and confusion, especially for the coders who are not trained as educational researchers. However, the results presented in Chapter three (Figure 5) offer a more "low-level" operationalization of the cognitive presence phases. For example, triggering event messages are short, tend to occur at the start of the discussion, have most question marks, have low lexical diversity, and mention few content-related concepts. They also tend to receive the most replies that have a little semantic resemblance to the originating triggering event message. This and similar descriptions provide a more detailed information about the important characteristics of each cognitive presence phase with regards to students' use of written language. By understanding the ways in which cognitive presence is expressed in student messages, not only do we provide a more detailed set of indicators of the different practical inquiry phases than currently available in the CoI literature, but more importantly, we improve the existing understanding of cognitive presence as a dynamic and cyclic process of practical inquiry.

8.1.3 RQ 3: Learning analytics assessment of cognitive presence

The central and most substantial part of the thesis is dedicated to developing learning analytics systems for assessing students' cognitive presence. In this regard, we developed two complementary systems which together provide a holistic assessment of cognitive presence within traditional online courses. The classification system presented in Chapter three provides an assessment approach to student cognitive presence as expressed in the discourse transcripts, while the clustering system presented in Chapter four provides an assessment of cognitive presence as captured by the trace data of students' interactions with a learning platform. Together, the two systems provide a holistic assessment of cognitive presence in both social (discourse-based) and personal (reflective) dimension of learning.

From the practical perspective, the most significant impact of the work presented in this thesis is the development of systems that can be used by the practitioners to assess students cognitive presence development without a need for manual, labor-intensive content analysis of discussion transcripts nor the use of self-report instruments. The classification system provides an assessment of cognitive presence on the level of each message, which enables identification of the parts of the discourse which require discussion facilitation and direct instruction by the teachers. In contrast, the clustering system provides an assessment of students' cognitive presence on the student level, which can be used to identify students that may require more instructional support regarding the use of available tools and resources. The clustering system described in this thesis is the first one to use data other than discussion transcripts to provide insights into students' cognitive presence development based on their individual, reflective learning activities, which – despite their critical significance – are often overlooked. Overall, the analytics developed in this thesis provide practical means for supporting the development of cognitive presence, which in turn, enable the broader adoption of the CoI model within different online and blended settings.

8.1.4 RQ 4: Evaluation of the Community of Inquiry model within Massive Open Online Courses

With the ability to process large amounts of educational data, one potential application of learning analytics is scaling up existing distance education pedagogies to novel online learning modalities where a sheer amount of interactions limits the ability of instructors to successfully monitor and support all students in the course. With this in mind, one of the key research questions in this thesis is whether we can develop successful methods for assessing students' cognitive presence within MOOC courses. However, before we could develop MOOC-specific learning analytics models for cognitive presence, we first needed to validate the applicability of the CoI model for capturing students' experience in MOOCs.

To compare MOOCs and traditional online courses regarding the CoI model, we examined the differences with regards to the reliability and validity of the CoI survey instrument (Arbaugh et al., 2008) between the two learning settings. The widespread adoption of the CoI survey instrument (Arbaugh et al., 2008) in traditional online courses resulted in a significant number of studies which examined the instrument's reliability and validity. In our study, we investigated the validity of the hypothesized, three-factor model of the CoI survey instrument, and the optimal model structure. As there are important differences between the traditional online courses and MOOCs (Ho et al., 2014; Kizilcec, Piech, & Schneider, 2013), the identified discrepancies and inconsistencies in the factor model structure between the two learning contexts would indicate the need for reevaluation of the CoI model and the survey instrument for assessing MOOC student experience. Our results confirmed that the CoI survey instrument is a reliable and valid tool for measuring students perceived levels of three presences and that the peculiarities of MOOC context do not undermine the theoretical foundation of the CoI model.

From the practical standpoint, the validation of the MOOC instrument within the MOOC context provides a justification for the use of the CoI instrument for assessing students' MOOC learning experience. Although the CoI instrument has been used in some MOOC studies (Poquet et al., 2017; Damm, 2016), to the best of our knowledge, our study is the first one to systematically evaluate the reliability and validity of the CoI instrument in the MOOC context. Moreover, given the disconnect between MOOC research and contemporary research in distance and online education (Rodriguez, 2012; Gašević et al., 2014; Storme, Vansieleghem, Devleminck, Masschelein, & Simons, 2016) our investigation provides a major step toward theoretically grounded MOOC research. Given raising criticism targeting MOOC pedagogies (Chamberlin & Parish, 2011; Durden, Bottomly, Rubin, Stetz, & Hirsch, 2013; Dolan, 2014), our investigation provides an approach to bridging between MOOC and traditional online research and a way to use the CoI model in a methodologically sound manner in MOOCs.

8.1.5 RQ 5: Cognitive presence assessment within Massive Open Online Courses

After the validation of the CoI model in the MOOC context, we investigated the use of trace data of students' use of learning tools and resources for the assessment of their cognitive presence. With the ability to process large quantities of data in an automated manner, learning analytics provide an outstanding opportunity to support richer and more intense course interactions by identifying a subset of students that require instructional support the most. In this regard, we developed a learning analytics system that can be used to reveal different student learning strategies and examined their association with the differences in student cognitive presence development. Our results revealed differences with respect to the students' level of cognitive presence and in particular, the resolution phase. Moreover, our study also identified important differences with respect to student motivation, goal orientation, and levels of self-regulation.

There are several significant contributions of the present work. First, our study revealed that a majority of students in MOOCs fail to regulate effectively their use of the available tools and resources. Although the same has been observed in traditional online courses (Clarebout, Elen, Collazo, Lust, & Jiang, 2013; Lust, Juarez Collazo, Elen, & Clarebout, 2012; Lust, Elen, & Clarebout, 2013), this problem is additionally emphasized in MOOCs, where the higher diversity of learners, massive student cohorts, and the need for strong learner autonomy seem to only negatively contribute to the quality of student experience (Kop, 2011; Guàrdia, Maina, & Sangrà, 2013; Hew & Cheung, 2014; Terras & Ramsay, 2015; Touati, 2016). Secondly, our results point out to the need for specific instructional support for students with different student learning strategies which cater to their various needs and enrollment motivations. For instance, disengaged students might benefit more from interventions that promote active use of online discussions rather than graded assignments, given their lack of interest in obtaining course certificates. Thirdly, our study also shows significant challenges regarding the use of survey data to assess baseline cognitive presence levels, given the low discriminatory power of the five-item Likert-scale. The likely causes for the low discriminatory power are self-selection bias, and also the different standards of participation by which students judge their own participation. Finally, as we closely followed the work presented in Chapter four, our results show significant promise of using and adjusting analytical models developed for traditional online courses in the MOOC setting.

8.2 Directions for future work

There are many promising directions for future work to expand the findings of this thesis. Generally speaking, those include work on additional forms of learning analytics relating to cognitive presence and analytics focusing on the other two dimensions (social and teaching presences) of the CoI model.

Given the importance of learning contexts on learning analytics systems (Gašević, Dawson, Rogers, & Gasevic, 2016), an important direction for future work is an examination of the extent to which learning context affects the accuracy of the classification system presented in Chapter three. While the classifier was developed using the data from a graduate level software engineering course with carefully designed discussions, it might be the case that the classification system would need a significant adjustment to work effectively in, for example, an undergraduate sociology class.

In addition to understanding the role of the course context for the accuracy of the classifier, an important area of future work relates to the role of the course context on the identified strategies of the use of learning tools. In Chapter four, we identified six different strategies. In future work, it would be important to investigate to what extent the role of the course design and context play a role in shaping up the strategies that a student adopts in a course. In Chapter seven, we see that strategies identified in the MOOC context were slightly different from the ones discovered in the traditional online setting (Kovanović, Joksimović, Poquet, Hennis, de Vries, et al., 2017). In the same way, we expect to find differences in the student learning strategies in the courses from other domains and with different course organization and structure.

Another important direction of future work revolves around the integration of the analytics presented in this thesis into existing learning platforms for the use by course instructors. In particular, it is critical to understand how the analytics developed in this thesis would be adopted by educational practitioners on a day-to-day basis. This would provide insights into the limitations and areas for improvement of the analytics presented in this thesis. Integration of those analytics systems into the learning platform would involve the design and implementation of a highly usable graphical interface for conducting analyses, interpreting analysis results, and ultimately, intervening on the basis of the conducted analyses. From a practical perspective, such interface would enable instructional interventions that are both data-informed and theoretically founded in the CoI model. One of the key challenges of learning analytics systems commonly available in the literature and practice is that they typically lack a theoretical foundation in established educational research (Gašević, Dawson, & Siemens, 2015). Therefore, the analytics and their graphical interface would provide an important example of a pedagogically-sound learning analytics implementation.

Given the critical importance of timely feedback on student learning success (Hattie & Timperley, 2007; Butler & Winne, 1995), an important area of future works is an investigation of the use of learning analytics for feedback provision. For example, the classification system could be integrated into online discussions so that immediate classification of a newly posted message provides feedback to students regarding the cognitive presence observed in their posting. For instance, a learning platform might provide a message notification to a student in cases where a non-cognitive message is being posted prompting the student to reconsider his contribution to the discussions [1] A similar classification system focusing on the message perceived quality by Weimer, Gurevych, and Mühlhäuser (2007) has already shown promise, and we believe that the same approach would be beneficial for developing students' cognitive presence. Next, the learning platform could also provide students with automated feedback regarding their use of the learning platform, with a list of suggestions that would improve their learning strategies.

With respect to the validation of the CoI model within a MOOC setting, an important direction for future work is an investigation of the relationship between three presences in a manner similar to the work of Shea and Bidjerano (2009) and Garrison, Cleveland-Innes, and Fung (2010). Given the specifics of MOOC learning and teaching, it is important to examine whether the causal relationship identified within the context of traditional online courses still holds or whether there are some peculiarities surrounding their interplay in the MOOC setting.

[1] It should be acknowledged that non-cognitive messages also have their place in student discussions, given their role in building student relationships and overall community development (Kenny, 2008).

Finally, the foundation of the CoI model is a holistic view of the student learning experience which is shaped by the productive and fruitful interaction of the three presences. As indicated by Garrison and Cleveland-Innes (2005) *"a deep approach to learning must consider all three elements of the community of inquiry: social, cognitive, and teaching presence."* (p. 144) As such, an important direction of future work is to extend the analytics developed in this thesis to cover both teaching and social presences, as a way of providing a comprehensive assessment of student online experience and learning outcomes. This would involve the development of classification systems for teaching and social presence based on online discussion transcripts, as well as the examination of how their levels are related to students' self-regulation, self-efficacy, metacognition, and goal orientation. Similarly, the assessment model presented in Chapter two would be extended to include teaching and social presence in a coherent assessment model of student online learning experience.

8.3 Conclusions

The underlying premise of the work presented in this thesis is that advanced analytics techniques and vast amounts of digital data and offer great potential for improving students' online learning experience and outcomes (Siemens et al., 2011). The focus of this thesis was on using the digital trace data for the assessment of cognitive presence, given its association with higher-order and critical thinking, which are – according to Garrison, Anderson, and Archer (2010) – *"the ultimate goal of higher education."* (p. 6) In particular, we focused on developing learning analytics based on the widely-adopted Community of Inquiry (CoI) model (Garrison et al., 1999), and its cognitive presence construct which captures the development of students' deep and critical thinking skills (Garrison et al., 2001). Given the challenges with the existing approaches to assessing cognitive presence by Garrison et al. (2001) and Arbaugh et al. (2008), the goal of this thesis was to utilize significant amounts of data readily available in learning environments for the development of learning analytics systems that can assess cognitive presence in an automated manner. Those analytics also provide novel theoretical insights into the nature of cognitive presence and its dynamics in a data-informed way, leading the path to better understanding of the cognitive presence construct and the underlying four phases of inquiry-based learning.

In this thesis, we described several systems for cognitive presence assessment which together provide an in-depth view of the development of cognitive presence, both in traditional online and MOOC contexts. The ultimate goal of those analytics is to enable the provision of timely information about students' learning progress which can be used to drive instructional interventions and provision of formative and actionable student feedback (Butler & Winne, 1995; Hattie & Timperley, 2007; Akyol & Garrison, 2011a). This is especially important for the adoption of CoI model in new modalities of online learning – such as MOOCs – where excessively large numbers of students make it challenging for instructors to monitor and support students in their learning activities. Finally, the work presented in this thesis provides one example how new developments in data science and analytics on the one hand, and contemporary educational research can be combined in a productive manner to tackle an important problem of the assessment of higher order learning.

Bibliography

Adams Becker, S., Cummins, M., Davis, A., Freeman, A., Hall Giesinger, C., & Ananthanarayanan, V. (2017). *NMC Horizon Report: 2017 Higher Education Edition*. The New Media Consortium. Austin, TX. (Cited on p. 1).

Aggarwal, C. C. & Zhai, C. (2012). *Mining Text Data*. Springer. (Cited on pp. 5, 50).

Akyol, Z., Arbaugh, J. B., Cleveland-Innes, M., Garrison, D. R., Ice, P., Richardson, J. C., & Swan, K. (2009). A Response to the Review of the Community of Inquiry Framework. *Journal of Distance Education*, *23*(2), 123–135. (Cited on p. 141).

Akyol, Z. & Garrison, D. R. (2011a). Assessing metacognition in an online community of inquiry. *The Internet and Higher Education*, *14*(3), 183–190. doi:10.1016/j.iheduc.2011.01.005. (Cited on pp. 65, 66, 84, 125, 174, 181)

Akyol, Z. & Garrison, D. R. (2011b). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology*, *42*(2), 233–250. doi:10.1111/j.1467-8535.2009.01029.x. (Cited on p. 2)

Anderson, T. & Dron, J. (2010). Three generations of distance education pedagogy. *The International Review of Research in Open and Distance Learning*, *12*(3), 80–97. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/890/. (Cited on pp. 10, 50, 125)

Anderson, T., Rourke, L., Garrison, D. R., & Archer, W. (2001). Assessing Teaching Presence in a Computer Conferencing Context. *Journal of Asynchronous Learning Networks*, *5*, 1–17. Retrieved from http://onlinelearningconsortium.org/wp-content/plugins/olc-%20download/download.php?post_id=4190. (Cited on pp. 2, 125)

Arbaugh, J., Cleveland-Innes, M., Diaz, S. R., Garrison, D. R., Ice, P., Richardson, J. C., & Swan, K. P. (2008). Developing a community of inquiry instrument: Testing a measure of the Community of Inquiry framework using a multi-institutional sample. *The Internet and Higher Education*, *11*(3–4), 133–136. doi:10.1016/j.iheduc.2008.06.003. (Cited on pp. 2, 4, 10, 125, 126, 141, 144, 178, 181)

Azevedo, R. & Aleven, V. (2013). Metacognition and Learning Technologies: An Overview of Current Interdisciplinary Research. In R. Azevedo & V. Aleven (Eds.), *International Handbook of*

Metacognition and Learning Technologies (28, pp. 1–16). Springer International Handbooks of Education. Springer New York. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-5546-%203_1. (Cited on pp. 65, 66)

Bali, M. (2014). MOOC Pedagogy: Gleaning Good Practice from Existing MOOCs. *Journal of Online Learning and Teaching*, *10*(1), 44–56. Retrieved from http://jolt.merlot.org/vol10no1/bali_0314.pdf. (Cited on p. 143)

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. doi:10.1037/0033-295X.84.2.191. (Cited on p. 65)

Bates, T. (2012). 9 What's right and what's wrong about Coursera-style MOOCs | Tony Bates. Retrieved from https://www.tonybates.ca/2012/08/05/whats-right-and-whats-%20wrong-about-coursera-style-moocs/. (Cited on p. 143)

Bloom, B. S. (1974). Time and learning. *American Psychologist*, *29*(9), 682–688. doi:10.1037/h0037632. (Cited on pp. 8, 86, 124)

Bond, L. (2014). A Brief Note on Evidence-Centered Design as a Mechanism for Assessment Development and Evaluation. *Measurement: Interdisciplinary Research and Perspectives*, *12*(1-2), 37–38. doi:10.1080/15366367.2014.921486. (Cited on p. 13)

Boxall, M. (2012). MOOCs: A massive opportunity for higher education, or digital hype? *The Guardian*. Retrieved from http://www.theguardian.com/higher-education-network/blog/%202012/aug/08/mooc-coursera-higher-education-investment. (Cited on p. 143)

Butler, D. L. & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245–281. doi:10.3102/00346543065003245. (Cited on pp. 66, 180, 181)

Carroll, J. (1963). A Model of School Learning. *Teachers College Record*, *64*(8), 723–733. (Cited on pp. 8, 86).

Chamberlin, L. & Parish, T. (2011). MOOCs: Massive Open Online Courses or Massive and Often Obtuse Courses? *eLearn*, *2011*(8). doi:10.1145/2016016.2016017. (Cited on p. 178)

Clarebout, G., Elen, J., Collazo, N. A. J., Lust, G., & Jiang, L. (2013). Metacognition and the Use of Tools. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (28, pp. 187–195). Springer International Handbooks of Education. Springer New York. doi:10.1007/978-1-4419-5546-3_13. (Cited on p. 179)

Clow, D. (2013). MOOCs and the Funnel of Participation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 185–189). LAK '13. New York, NY, USA: ACM. doi:10.1145/2460296.2460332. (Cited on p. 144)

Corich, S., Hunt, K., & Hunt, L. (2012). Computerised Content Analysis for Measuring Critical Thinking within Discussion Forums. *Journal of e-Learning and Knowledge Society*, *2*(1), 47–60. Retrieved from http://www.je-lks.org/ojs/index.php/Je-LKS_EN/article/view/%20700. (Cited on pp. 7, 62)

Damm, C. A. V. (2016). Applying a Community of Inquiry Instrument to Measure Student Engagement in Large Online Courses. *Current Issues in Emerging eLearning*, *3*(1), 138–172. Retrieved from http://scholarworks.umb.edu/ciee/vol3/iss1/9/. (Cited on p. 178)

Daniel, J. (2014). Foreword to the Special Section on Massive Open Online Courses MOOCs – Evolution or Revolution? *Journal of Online Learning and Teaching*, *10*(1), i–iv. Retrieved from http://jolt.merlot.org/vol10no1/daniel_foreword_0314.pdf. (Cited on p. 144)

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*, *46*(1), 6–28. (Cited on p. 48).

Dewey, J. (1910). *How We Think*. D.C. Heath & Company. (Cited on pp. 1, 48, 49).

Dolan, V. L. B. (2014). Massive online obsessive compulsion: What are they saying out there about the latest phenomenon in higher education? *The International Review of Research in Open and Distributed Learning*, *15*(2). Retrieved May 16, 2017, from http://www.irrodl.org/index.php/irrodl/article/view/1553. (Cited on p. 178)

Donnelly, R. & Gardner, J. (2011). Content analysis of computer conferencing transcripts. *Interactive Learning Environments*, *19*(4), 303–315. (Cited on p. 49).

Durden, W. G., Bottomly, H. K., Rubin, B., Stetz, M. D., & Hirsch, M. (2013). Online Courses: Possibilities and Pitfalls. *The New York Times*. Retrieved May 16, 2017, from http://www.nytimes.com/2013/01/29/opinion/online-courses-possibilities-and-pitfalls.html. (Cited on p. 178)

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Los Angeles; London: SAGE. (Cited on p. 5).

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. doi:10.1037/0003-066X.34.10.906. (Cited on pp. 65, 66)

Garrison, D. R. (1993). A cognitive constructivist view of distance education: An analysis of teaching-learning assumptions. *Distance Education*, *14*(2), 199–211. doi:10.1080/0158791930140204. (Cited on p. 10)

Garrison, D. R. (2011). Social Presence. In *E-Learning in the 21st Century: A Framework for Research and Practice* (pp. 30–41). Taylor & Francis. (Cited on p. 143).

Garrison, D. R. & Akyol, Z. (2013). Toward the development of a metacognition construct for communities of inquiry. *The Internet and Higher Education*, *17*, 84–89. doi:10.1016/j.iheduc.2012.11.005. (Cited on pp. 65, 66, 174)

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education*, *2*(2–3), 87–105. doi:10.1016/S1096-7516(00)00016-6. (Cited on pp. 1, 2, 10, 49, 125, 181)

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education. *American Journal of Distance Education*, *15*(1), 7–

23. doi:10.1080/08923640109527071. (Cited on pp. 2, 4, 5, 7, 12, 13, 47, 49, 62, 64, 84, 174, 176, 181)

Garrison, D. R., Anderson, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education*, *13*(1–2), 5–9. doi:10.1016/j.iheduc.2009.10.003. (Cited on pp. 49, 125, 181)

Garrison, D. R. & Arbaugh, J. (2007). Researching the community of inquiry framework: Review, issues, and future directions. *The Internet and Higher Education*, *10*(3), 157–172. doi:10.1016/j.iheduc.2007.04.001. (Cited on p. 125)

Garrison, D. R. & Archer, W. (2000). *A Transactional Perspective on Teaching and Learning: A Framework for Adult and Higher Education*. Amsterdam ; New York: Elsevier Science. (Cited on p. 7).

Garrison, D. R. & Cleveland-Innes, M. (2005). Facilitating Cognitive Presence in Online Learning: Interaction Is Not Enough. *American Journal of Distance Education*, *19*(3), 133–148. doi:10.1207/s15389286ajde1903_2. (Cited on p. 181)

Garrison, D. R., Cleveland-Innes, M., & Fung, T. S. (2010). Exploring causal relationships among teaching, cognitive and social presence: Student perceptions of the community of inquiry framework. *The Internet and Higher Education*, *13*(1–2), 31–36. doi:10.1016/j.iheduc.2009.10.002. (Cited on pp. 125, 180)

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. doi:10.1016/j.iheduc.2015.10.002. (Cited on p. 179)

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. doi:10.1007/s11528-014-0822-x. (Cited on p. 180)

Gašević, D., Kovanović, V., Joksimović, S., & Siemens, G. (2014). Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative. *The International Review of Research in Open and Distributed Learning*, *15*(5). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1954. (Cited on pp. 143, 178)

Gorsky, P., Caspi, A., Blau, I., Vine, Y., & Billet, A. (2011). Toward a CoI population parameter: The impact of unit (sentence vs. message) on the results of quantitative content analysis. *The International Review of Research in Open and Distance Learning*, *13*(1), 17–37. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1073. (Cited on pp. 125, 141)

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix Providing Multilevel Analyses of Text Characteristics. *Educational Researcher*, *40*(5), 223–234. doi:10.3102/0013189X11413260. (Cited on p. 50)

Guàrdia, L., Maina, M., & Sangrà, A. (2013). MOOC design principles: A pedagogical approach from the learner's perspective. *eLearning Papers*, (33). Retrieved from http://www.openeducationeuropa.eu/en/article/MOOC-Design-Principles.-A-Pedagogical-Approach-from-the-Learner%E2%80%99s-Perspective?paper=124335. (Cited on p. 179)

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*(2-3), 107–124. doi:10.1007/s11409-007-9016-7. (Cited on pp. 64, 66)

Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. doi:10.3102/003465430298487. (Cited on pp. 180, 181)

Henri, F. (1992). Computer Conferencing and Content Analysis. In A. R. Kaye (Ed.), *Collaborative Learning Through Computer Conferencing* (90, pp. 117–136). NATO ASI Series. Springer Berlin Heidelberg. (Cited on p. 48).

Hew, K. F. & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review*, *12*, 45–58. doi:10.1016/j.edurev.2014.05.001. (Cited on p. 179)

Ho, A. D., Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). *HarvardX and MITx: The First Year of Open Online Courses, Fall 2012-Summer 2013* (SSRN Scholarly Paper No. ID 2381263). Social Science Research Network. Rochester, NY. Retrieved from http://papers.ssrn.com/abstract=2381263. (Cited on pp. 144, 178)

Jézégou, A. (2010). Community of Inquiry in E-learning: A Critical Analysis of the Garrison and Anderson Model (Winner 2010 Best Research Article Award). *International Journal of E-Learning & Distance Education*, *24*(3). Retrieved from http://www.ijede.ca/index.php/jde/article/view/707. (Cited on pp. 2, 125)

Joksimović, S., Skrypnyk, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., Dawson, S., Graesser, A. C., & Brooks, C. (2017). How do we model learning at scale? A systematic review of the literature. *Manuscript submitted for publication*. (Cited on p. 143).

Kalz, M. & Specht, M. (2013). If MOOCS are the answer, did we ask the right questions? Implications for the design of large-scale online-courses. In *Proceedings of the MSM's 3rd Annual Research Conference*. Maastricht School of Management. Retrieved from http://dspace.ou.nl/handle/1820/5183. (Cited on p. 143)

Karweit, N. (1982). *Time-on-task: A research review* (tech. rep. No. 332). Johns Hopkins University. Baltimore, MD, United States. (Cited on p. 124).

Karweit, N. (1984). Time-on-Task Reconsidered: Synthesis of Research on Time and Learning. *Educational Leadership*, *41*(8), 32–35. (Cited on p. 124).

Karweit, N. & Slavin, R. E. (1982). Time-on-task: Issues of timing, sampling, and definition. *Journal of Educational Psychology*, *74*(6), 844–851. doi:10.1037/0022-0663.74.6.844. (Cited on p. 124)

Kenny, M. A. (2008). Discussion, Cooperation, Collaboration. *The Interpreter and Translator Trainer*, *2*(2), 139–164. doi:10.1080/1750399X.2008.10798771. (Cited on p. 180)

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses. In *Proceedings of the Third International Con-*

ference on Learning Analytics and Knowledge (pp. 170–179). LAK '13. New York, NY, USA: ACM. doi:10.1145/2460296.2460330. (Cited on p. 178)

Kizilcec, R. F. & Schneider, E. (2015). Motivation As a Lens to Understand Online Learners: Toward Data-Driven Design with the OLEI Scale. *ACM Transactions on Computer-Human Interaction*, *22*(2), 6:1–6:24. doi:10.1145/2699735. (Cited on p. 144)

Kop, R. (2011). The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course. *The International Review of Research in Open and Distance Learning*, *12*(3), 19–38. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/882. (Cited on p. 179)

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., & Baker, R. S. (2015). Does time-on-task estimation matter? Implications on validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. doi:10.18608/jla.2015.23.6. (Cited on pp. viii, 9, 87)

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015). Penetrating the black box of time-on-task estimation. In *Proceedings of the Fifth International Conference on Learning Analytics & Knowledge (LAK'15)* (pp. 184–193). LAK '15. (best paper award winner). New York, NY, USA: ACM. doi:10.1145/2723576.2723623. (Cited on p. 9)

Kovanović, V., Gašević, D., & Hatala, M. (2014). Learning analytics for communities of inquiry. *Journal of Learning Analytics*, *1*(3), 195–198. Retrieved from http://epress.lib.uts.edu.au/journals/index.php/JLA/article/%20view/4234. (Cited on p. 2)

Kovanović, V., Gašević, D., Hatala, M., & Siemens, G. (2017). *A novel model of cognitive presence assessment using automated learning analytics methods*. SRI Education. Retrieved from http://a4li.sri.com/archive/papers/Kovanovic_2017_Presence.pdf. (Cited on pp. viii, 6, 13, 62, 84)

Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, *27*, 74–89. doi:10.1016/j.iheduc.2015.06.002. (Cited on pp. viii, 8, 11, 67, 84)

Kovanović, V., Joksimović, S., Gašević, D., & Hatala, M. (2014). Automated Cognitive Presence Detection in Online Discussion Transcripts. In *Proceedings of the Workshops at the LAK 2014 Conference co-located with 4th International Conference on Learning Analytics and Knowledge (LAK'14)*. Indianapolis, IN. Retrieved from http://ceur-ws.org/Vol-1137/. (Cited on pp. 7, 62)

Kovanović, V., Joksimović, S., Gašević, D., Siemens, G., & Hatala, M. (2015a). MOOCs in the news: A European perspective. In *Proceedings of the 2015 HOME Conference* (pp. 36–48). Rome, Italy: EADTU. Retrieved from http://eadtu.eu/home/publications. (Cited on p. 143)

Kovanović, V., Joksimović, S., Gašević, D., Siemens, G., & Hatala, M. (2015b). What public media reveals about MOOCs: A systematic analysis of news reports. *British Journal of Educational Technology*, *46*(3), 510–527. doi:10.1111/bjet.12277. (Cited on p. 143)

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., Čukić, I., de Vries, P., Hatala, M., Dawson, S., Siemens, G., & Gašević, D. (2017). Exploring communities of inquiry in massive open online courses. *Manuscript submitted for publication*. (Cited on pp. viii, 10, 126).

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., de Vries, P., Hatala, M., Dawson, S., Siemens, G., & Gašević, D. (2017). The role of technology use on shaping student learning experience in MOOCs. *Manuscript submitted for publication*. (Cited on pp. viii, 11, 145, 180).

Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK'16)* (pp. 15–24). LAK '16. New York, NY, USA: ACM. doi:10.1145/2883851.2883950. (Cited on pp. viii, 7, 51, 62)

Krippendorff, K. H. (2003). *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage Publications. (Cited on pp. 2, 5, 49).

Larkin, S. (2009). Socially mediated metacognition and learning to write. *Thinking Skills and Creativity*. Applying Thinking Skills within Educational Settings and Beyond, *4*(3), 149–159. doi:10.1016/j.tsc.2009.09.003. (Cited on pp. 65, 143)

Lipman, M. (1991). *Thinking in education*. Cambridge University Press. (Cited on pp. 1, 48, 49).

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment. *ETS Research Report Series*, *2014*(1), 1–23. doi:10.1002/ets2.12009. (Cited on p. 1)

Luppicini, R. (2007). Review of computer mediated communication research for education. *Instructional Science*, *35*(2), 141–185. doi:10.1007/s11251-006-9001-6. (Cited on p. 48)

Lust, G., Elen, J., & Clarebout, G. (2013). Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. *Computers in Human Behavior*, *29*(5), 2013–2021. doi:10.1016/j.chb.2013.03.014. (Cited on p. 179)

Lust, G., Juarez Collazo, N. A., Elen, J., & Clarebout, G. (2012). Content Management Systems: Enriched learning opportunities for all? *Computers in Human Behavior*, *28*(3), 795–808. doi:10.1016/j.chb.2011.12.009. (Cited on p. 179)

Manning, C. D., Schütze, H. et al. (1999). *Foundations of statistical natural language processing*. MIT Press. (Cited on p. 50).

Martinez, M. E. (2006). What is metacognition? *Phi delta kappan*, *87*(9), 696–699. (Cited on p. 65).

McKlin, T. (2004). *Analyzing Cognitive Presence in Online Courses Using an Artificial Neural Network* (Doctoral dissertation, Georgia State University, College of Education, Atlanta, GA, United States). (Cited on pp. 7, 62).

McKlin, T., Harmon, S., Evans, W., & Jones, M. (2002). Cognitive presence in web-based learning: A content analysis of students' online discussions. In *IT Forum* (Vol. 60). (Cited on p. 62).

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A Brief Introduction to Evidence-Centered Design. *ETS Research Report Series*, *2003*(1), i–29. doi:10.1002/j.2333-8504.2003.tb01908.x. (Cited on pp. 5, 12)

Mislevy, R. J., Behrens, J. T., DiCerbo, K. E., & Levy, R. (2012). Design and Discovery in Educational Assessment: Evidence-Centered Design, Psychometrics, and Educational Data Mining. *Journal of Educational Data Mining*, *4*(1), 11–48. Retrieved from http://researchnetwork.pearson.com/wp-content/uploads/%20MislevyEtAlVol4Issue1P11_48.pdf. (Cited on p. 13)

Mislevy, R. J. & Haertel, G. D. (2006). Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice*, *25*(4), 6–20. doi:10.1111/j.1745-3992.2006.00075.x. (Cited on p. 12)

OCR. (2015). The 50 Most Popular MOOCs of All Time. Retrieved from http://www.onlinecoursereport.com/the-50-most-popular-moocs-%20of-all-time/. (Cited on p. 126)

Pappano, L. (2012). The Year of the MOOC. *New York Times*. Retrieved from http://www.nytimes.com/2012/11/04/education/edlife/massive-%20open-online-courses-are-multiplying-at-a-rapid-pace.html?%20pagewanted=all. (Cited on p. 143)

Pintrich, P. R. & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. doi:10.1037/0022-0663.82.1.33. (Cited on p. 65)

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (Mslq). *Educational and Psychological Measurement*, *53*(3), 801–813. doi:10.1177/0013164493053003024. (Cited on p. 66)

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)* (tech. rep. No. 91-B-004). University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning. Ann Arbor, MI. (Cited on p. 66).

Poquet, O., Kovanović, V., de Vries, P., Hennis, T., Joksimović, S., Gašević, D., & Dawson, S. (2017). Social presence in massive open online courses. *Manuscript submitted for publication*. (Cited on p. 178).

Reich, J. (2015). Rebooting MOOC Research. *Science*, *347*(6217), 34–35. doi:10.1126/science.1261627. (Cited on p. 143)

Riffe, D. & Freitag, A. (1997). A Content Analysis of Content Analyses: Twenty-Five Years of Journalism Quarterly. *Journalism & Mass Communication Quarterly*, *74*(3), 515–524. doi:10.1177/107769909707400306. (Cited on p. 49)

Rodriguez, C. O. (2012). MOOCs and the AI-Stanford like Courses: Two Successful and Distinct Course Formats for Massive Open Online Courses. *European Journal of Open, Distance and E-Learning*, *2012*(2), 1–13. (Cited on pp. 143, 178).

Ross, J., Sinclair, C., Knox, J., Bayne, S., & Macleod, H. (2014). Teacher Experiences and Academic Identity: The Missing Components of MOOC Pedagogy. *Journal of Online Learning and Teaching*, *10*(1), 57–69. (Cited on p. 144).

Rourke, L. (2000). *Exploring Social Communication in Cornputer Conferencing* (Doctoral dissertation, University of Alberta, Faculty of Graduate Studies and Research, Edmonton, AB, Canada). (Cited on p. 143).

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing Social Presence In Asynchronous Text-based Computer Conferencing. *The Journal of Distance Education*, *14*(2), 50–71. Retrieved from http://www.ijede.ca/index.php/jde/article/view/153. (Cited on p. 2)

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological Issues in the Content Analysis of Computer Conference Transcripts. *International Journal of Artificial Intelligence in Education (IJAIED)*, *12*, 8–22. (Cited on pp. 2, 49, 176).

Sandeen, C. (2013). Assessment's Place in the New MOOC World. *Research & Practice in Assessment; Lynchburg*, *8*. Retrieved from http://search.proquest.com/docview/1505317186/abstract/%203BBFECE02C9C48D7PQ/1. (Cited on p. 144)

Shah, D. (2017). MOOCs Started Out Completely Free. Where Are They Now? *EdSurge*. Retrieved from https://www.edsurge.com/news/2017-04-20-moocs-started-out-%20completely-free-where-are-they-now. (Cited on p. 144)

Sharma, P. & Hannafin, M. (2004). Scaffolding Critical Thinking in an Online Course: An Exploratory Study. *Journal of Educational Computing Research*, *31*(2), 181–208. doi:10.2190/TMC3-RXPE-75MY-31YG. (Cited on p. 65)

Shea, P. & Bidjerano, T. (2009). Community of inquiry as a theoretical framework to foster "epistemic engagement" and "cognitive presence" in online education. *Computers & Education*, *52*(3), 543–553. doi:10.1016/j.compedu.2008.10.007. (Cited on p. 180)

Shea, P. & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computers & Education*, *55*(4), 1721–1731. doi:10.1016/j.compedu.2010.07.017. (Cited on pp. 2, 65, 66, 125, 174)

Shea, P. & Bidjerano, T. (2012). Learning presence as a moderator in the community of inquiry model. *Computers & Education*, *59*(2), 316–326. doi:10.1016/j.compedu.2012.01.011. (Cited on pp. 65, 66, 174)

Shea, P., Hayes, S., Smith, S. U., Vickers, J., Bidjerano, T., Pickett, A., Gozza-Cohen, M., Wilde, J., & Jian, S. (2012). Learning presence: Additional research on a new conceptual element within the Community of Inquiry (CoI) framework. *The Internet and Higher Education*. Special issue of the American Educational Research Association's online teaching and learning special interest group, *15*(2), 89–95. doi:10.1016/j.iheduc.2011.08.002. (Cited on pp. 65, 66, 174)

Siemens, G. (2005). Connectivism: A learning theory for the digital age. In *International Journal of Instructional Technology and Distance Learning*. (Cited on p. 144).

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S., Ferguson, R., Duval, E., Verbert, K., & Baker, R. S. J. d. (2011). *Open Learning Analytics: An integrated & modularized platform*. SoLAR. (Cited on pp. 86, 181).

Snow, E., Haertel, G., Fulkerson, D., Feng, M., & Nichols, P. (2010). Leveraging Evidence-Centered Assessment Design in Large-Scale and Formative Assessment Practices. In *Proceedings of the 2010 Annual Meeting of the National Council on Measurement in Education (NCME)*. (Cited on p. 13).

Stacey, P. (2013). The pedagogy of MOOCs. *The International Journal for Innovation and Quality in Learning*, *11*(3), 111–115. Retrieved from http://empower.eadtu.eu/images/fields-of-expertise/%20OERsMOOCs/INNOQUAL-Issue-3-Publication-Sep-2014-FINAL-w-%20cover.pdf. (Cited on p. 143)

Stallings, J. (1980). Allocated Academic Learning Time Revisited, or Beyond Time on Task. *Educational Researcher*, *9*(11), 11–16. doi:10.3102/0013189X009011011. (Cited on p. 124)

Storme, T., Vansieleghem, N., Devleminck, S., Masschelein, J., & Simons, M. (2016). The emerging pedagogy of MOOCs, the educational design of technology and practices of study. *Journal of Computers in Education*, *3*(3), 309–328. doi:10.1007/s40692-016-0070-5. (Cited on p. 178)

Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education*, *46*(1), 29–48. (Cited on p. 49).

Tausczik, Y. R. & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. doi:10.1177/0261927X09351676. (Cited on p. 50)

Terras, M. M. & Ramsay, J. (2015). Massive open online courses (MOOCs): Insights and challenges from a psychological perspective. *British Journal of Educational Technology*, *46*(3), 472–487. doi:10.1111/bjet.12274. (Cited on p. 179)

Touati, A. (2016). Self-Directed Learning in MOOCs: A Disconnect Between Theory and Practice. *MAJER Middle Eastern & African Journal of Educational Research*, (19), 15–30. Retrieved from http://www.majersite.org/issue19/19_2.pdf. (Cited on p. 179)

Trilling, B. & Fadel, C. (2009). *21st century skills: Learning for life in our times*. John Wiley & Sons. (Cited on p. 1).

Waters, Z., Kovanović, V., Kitto, K., & Gašević, D. (2015). Structure matters: Adoption of structured classification approach in the context of cognitive presence classification. In G. Zuccon, S. Geva, H. Joho, F. Scholer, A. Sun, & P. Zhang (Eds.), *Information Retrieval Technology* (9460, pp. 227–238). Lecture Notes in Computer Science. Springer International Publishing. doi:10.1007/978-3-319-28940-3_18. (Cited on p. 62)

Weimer, M., Gurevych, I., & Mühlhäuser, M. (2007). Automatically Assessing the Post Quality in Online Discussions on Software. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions* (pp. 125–128). ACL '07. Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1557769.1557806. (Cited on p. 180)

Winne, P. H. (2006). How Software Technologies Can Improve Research on Learning and Bolster School Reform. *Educational Psychologist*, *41*(1), 5–17. doi:10.1207/s15326985ep4101_3. (Cited on pp. 64, 66)

Winne, P. H. & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). The educational psychology series. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. (Cited on p. 65).

Wise, A. F., Hausknecht, S. N., & Zhao, Y. (2014). Attending to others' posts in asynchronous discussions: Learners' online "listening" and its relationship to speaking. *International Journal of Computer-Supported Collaborative Learning*, *9*(2), 185–209. doi:10.1007/s11412-014-9192-9. (Cited on p. 84)

Wise, A. F., Speer, J., Marbouti, F., & Hsiao, Y.-T. (2013). Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviors. *Instructional Science*, *41*(2), 323–343. doi:10.1007/s11251-012-9230-9. (Cited on p. 84)

Zhou, M. & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, *22*(6), 413–419. doi:10.1016/j.learninstruc.2012.03.004. (Cited on p. 66)

Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*, *41*(2), 64–70. (Cited on pp. 65, 66).